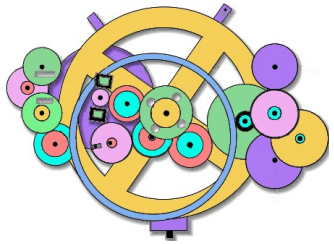
Antikythera Publications



Relational Database Design
http://www.AntikytheraPubs.com
foberle@AntikytheraPubs.com

IME Keyboard Layout Charts for selected Languages – "Hello, World"
Multi-script Database Series #8

Prepared by: Frank Oberle

Earlier Database Design Notes introduced many concepts that database administrators and developers should be familiar with in order to manage multi-lingual data. DDN #2 "Exploring Complex Text Layout," for example, introduced six common non-Latin scripts. To further assist with text entry, analysis, and troubleshooting the unique issues that can result in mixed-script environments encountered in business and academia, this note provides printable keyboard layout charts and reference tables for a variety of representative languages, using some common Input Method Engines (IMEs).

Just as writing a traditional "Hello World" application familiarizes us with the basic syntactic conventions of unfamiliar programming languages, examining some simple phrases in a variety of Scripts can serve as a high-level overview of the general characteristics and quirks of an unfamiliar writing system. While most developers will have no need or desire to read and write as many human languages as programming languages, having the ability to recognize a script and/or language can still be helpful – if only to determine where to seek help in resolving any issues that arise.

Revised July 2021 for public distribution
See pages 2 and 36 for information on other material from Antikythera Publications.

Database Design Note Series on maintaining Multi-Language/Multi-Script Databases

(All available for download from www.AntikytheraPubs.com/i18n.htm)

1. Exploring Alphabets
2. Exploring Complex Text Layout
3. Exploring UTF-8
4. Evaluating Fonts for use in Multi-Lingual Documents
5. Exploring Bidirectional (BIDI) Text Entry
6. Exploring Arabic Script Behavior
7. Exploring Han Script Entry – Chinese
8. Keyboard Layouts – Hello World
9. Evaluating Bidirectional Text Handling Behavior in Applications

Common IME Keyboard Layout Mappings for ANSI (U.S. English) Keyboards

Arabic | Chinese (Han) | Farsi | Greek x2 | Hebrew | Hindi | Japanese x2 | Korean | Mycenaean | Russian | Thai | Turkish | Urdu

There are two pages for each keyboard layout presented here; the first of these shows the mapping of each key on a U.S. English keyboard[CAUTION] to the character that will be produced by the IME when that key is pressed.

The title bar of each page is a loose translation of the phrase "*Language* Keyboard Layout" in the *language* which the layout illustrates. Below the title is its approximate English pronunciation, and the Latin keystrokes the IME requires to produce the name using the keyboard illustrated.

The space bar has a list of the Unicode planes from which the language derives its characters – a valuable source of information for each. As a further aid in identifying these, **Consonants are shown in Red**, **Vowels in Blue** (recognizing of course that, unlike Alphabets, Abjads have no vowels), **Numerals in Green**, and **Diacritics and Dead Keys in Purple**

The IME engine on which each Layout is based is shown on the left of the Right Shift key on each page. The direction of text for each language is also marked on the right side of that key as either LTR (text flows left to right, as it does in English or French) or RTL (right to left, as does in Arabic or Hebrew).

If you have used only LTR writing systems, consider that the process of moving the cursor forward through text is associated with the "right arrow." To successfully deal with bidirectional writing systems, you need to consider that the right arrow means "move *forward* through text" – not "move *right* through text"; likewise, the left arrow is "move *backward*" and the usual arrows displayed on the cursor keys as well as the Backspace key should be recognized as only applicable to languages using LTR writing systems.

Likewise the variety of paired symbols – e.g. ( ), [ ], and { } must be viewed as "opening/closing" rather than "left/right" symbols. This perspective permits their use on RTL keyboard layouts to be consistent with LTR mappings.

Many layout pages include a supplemental table showing similar characters side-by-side (with relevant information); this permits easier identification of such characters by those unfamiliar with the language/script being examined.

A second page for each language shows Unicode hexadecimal values – both shifted and unshifted – for each key on the keyboard layout. Some have extra tables sorted by hex values to make identifying the required keypresses easier.

In some cases, data stored on disk reflects a single composite character – not those actually entered, making it difficult to "reverse engineer" the key strokes needed to regenerate the original input text. Entering the Japanese Hiragana character か (t u+304b) followed by a ゛ (゛ u+309b) will cause the が composite (u+304c, for which no single keypress is available) to be stored on disk. Similarly, typing e / on a Mizuochi Ancient Greek layout will combine, display, and store a composite έ (u+1f73), not the separate epsilon and acute accent. Therefore, some layouts include additional tables to assist identifying the key sequences used to generate such composite characters.

The term "Chinese," discussed in detail in DDN-7, can refer to a wide variety of spoken languages but since written Han – the Hànzì glyphs used by all of these as well as other languages – is most commonly entered with an English layout, no such layout is given here. Some examples of Mandarin are provided however to aid in testing and troubleshooting Han data issues.

CAUTION: This set of keyboard maps assumes the use of an ANSI U.S. English (QWERTY) keyboard; if, for example, a Dvorak, AZERTY (used in France and Belgium) or QWERTZ (used in other central European countries) layout is used, these must be adapted accordingly. With a few caveats, however, this set of keyboard layouts are still applicable to ISO English (QWERTY) keyboards used outside the U.S. – generally identified by their larger Enter key.

IMEs for U.S. ANSI layouts often do not define keys for currency symbols other than the U.S. dollar ($); symbols for Euro (€), Rupee (₹), Yen (¥), Ruble (₽) and others cannot be directly entered, though two exceptions – the Iranian Rial (﷼) and Thai Baht (฿) – are among the layouts provided here.

Keyboard Layouts for Selected Scripts
Using Representative Languages & IMEs

Tables of Keyboard Mappings and Code Points

1 Interestingly, “Japanese Language” is written in Kanji – adapted in the distant past from Chinese Hànzì Script – rather than the native Hiragana.
2 No word for the concept of ‘language’ has been uncovered in any Linear-B writings in Knossos or the Greek mainland to date. The final “s” was inferred.

تخطيط لوحة المفاتيح العربية

Arabic (Modern Standard) Layout

Pronunciation: takhtit lawhat almafatih Al'Arabía

← ***IME Keystrokes***: jo'd' g,pm hglthjdp hguvfdm

Similar Glyphs / Letter Forms in the Right-to-Left Arabic Impure Abjad

N.B. Numeric Characters in RTL Languages are displayed in LTR order.

ى	ي	ئ	ي	ب	ت	ث	ح	ج	خ	س	ش	ص	ض	ط	ظ	ع	غ	ف	ق	ك	ل	ن	ه	ة	لا	لآ	لأ
0649	0620	0626	064a	0628	062a	062b	062d	062c	062e	0633	0634	0635	0636	0637	0638	0639	063a	0641	0642	0643	0644	0646	0647	0629	fefb	fef5	fef7
n		z	d	f	j	e	p	[	o	s	a	w	q	'	/	u	y	t	r	;	g	k	i	m	b	B	G

The Arabic abjad (أبجدية عربية) is used for a variety of languages, including Arwi (in Sri Lanka and Southern India), Azerbaijani (in Iran), Baloch, Balti, Belarusian (among Belarusian Tatars), Bosniaks (in Bosnia), Brahui (in Pakistan), Central Kurdish, Chinese (in some areas), Farsi/Persian*, Fulfulde-Pular, Hausa, Kashmiri, Kazakh and Kyrgyz (in Central Asia), Kurdish (in Iraq and Iran), Kyrgyz (Central Asia), Luri, Malay, Mozarabic, Ottoman (pre-1928) Turkish, Panjabi (in Pakistan), Pashto, Sindhi, dialects of Mandinka (in West Africa), Swahili (in East Africa), Urdu*, Uyghur (in China and Central Asia), and Uzbek (in Central Asia). Note the paired delimiter reversals on the [,], D, F, C, and V keys reflecting the open/close view rather than the left/right view. The key combination shift+9 on the RTL Arabic keyboard is interpreted as "open Parenthesis" just as it would be on an LTR English keyboard.

N.B. Some keyboard variants such as m17n mappings use Indo-Arabic Number forms: 1=١; 2=٢; 3=٣; 4=٤; 5=٥; 6=٦; 7=٧; 8=٨; 9=٩; 0=٠; these have Unicode values u+63f0 to u+63f9 for digits 0 through 9 respectively. Arabic characters are often displayed (but not stored) using different initial, medial and final forms.

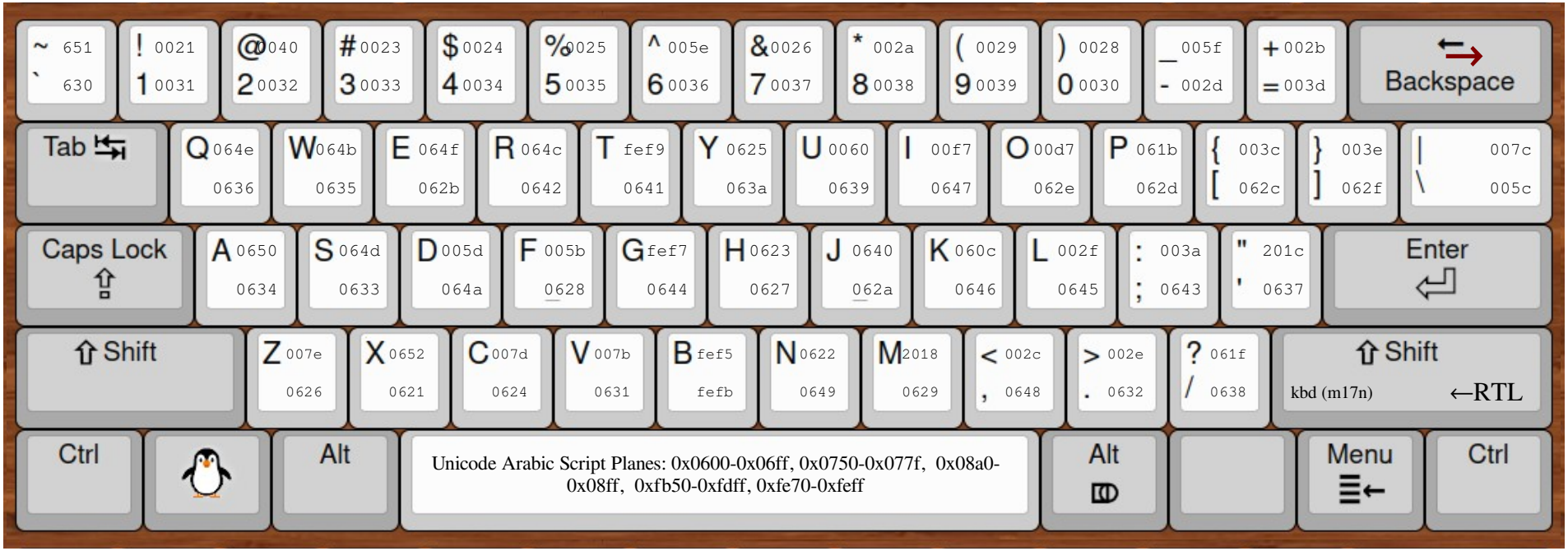
* See the Farsi/Iranian/Persian and Urdu keyboards on pages 7 and 31 respectively. Also see DDN-6 "Exploring Arabic Script Behavior" for more information on this complex Script.

تخطيط لوحة المفاتيح العربية

Arabic (Modern Standard) Code Points

Pronunciation: takhtit lawhat almafatih Al'Arabía

← IME Keystrokes: jo'd' g,pm hglthjdp hguvfdm



Modern Standard Arabic Text Analysis Example

To determine the key presses required to type لوحة (painting or writing) – the second word[α] of the title above – copy it into a word processor and move the cursor (▌) back and forth using the arrow keys, noting how, in RTL text, ← is interpreted as **forward** rather than **left**.

e.g. ▌لوحة ↔ ة▌لوح ↔ حة▌لو ↔ وحة▌ل ↔ لوحة▌

Note that there are 4 characters. Now place the cursor after (*to the left of*) the first character – " وحة▏ل ". Use the Word Processor command Alt+x[β], to view the character's hexadecimal Unicode value.

The text will change from " وحة▌ل " to "U+0644 ▌وحة"; the disconcerting placement of the Unicode value of "ل" from the far right of the word to its left is because the string "U+0644" consists of all Latin LTR characters.

Without repositioning the cursor, press the forward/right arrow to go past the remaining characters in sequence, pressing Alt+x[β] after each – the end result is:

U+0644U+0648U+062dU+0629 ▌ – with all codes in LTR sequence.

Look up each Unicode value in the table's K column: u+644 is a g, u+648 a comma, and so forth; therefore لوحة can be typed using the g , p m key sequence.

Unicode	UTF-8	K	Unicode	UTF-8	K	Unicode	UTF-8	K	Unicode	UTF-8	K	Unicode	UTF-8	K
u+60c	D88C	K	u+629	D8A9	m	u+634	D8B4	a	u+644	D8C4	g	u+64f	D98F	E
u+61b	D89B	P	u+62a	D8AA	j	u+635	D8B5	w	u+645	D8C5	l	u+650	D990	A
u+61f	D89F	?	u+62b	D8AB	e	u+636	D8B6	q	u+646	D8C6	k	u+651	D991	~
u+621	D8A1	x	u+62c	D8AC	[	u+637	D8B7	'	u+647	D8C7	i	u+0652	D991	X
u+622	D8A2	N	u+62d	D8AD	p	u+638	D8B8	/	u+648	D8C8	,	u+2018	E28098	M
u+623	D8A3	H	u+62e	D8AE	o	u+639	D8B9	u	u+649	D8C9	n	u+201c	E2809C	"
u+624	D8A4	c	u+62f	D8AF	]	u+63a	D8BA	y	u+64a	D98A	d			
u+625	D8A5	Y	u+630	D8B0	`	u+640	D8C0	J	u+64b	D98B	W	u+fef5	D984D8A2	B
u+626	D8A6	z	u+631	D8B1	v	u+641	D8C1	t	u+64c	D98C	R	u+fef7	D984D8A3	G
u+627	D8A7	h	u+632	D8B2	.	u+642	D8C2	r	u+64d	D98D	S	u+fef9	D984D8A5	T
u+628	D8A8	f	u+633	D8B3	s	u+643	D8C3	;	u+64e	D98E	Q	u+fefb	D984D8A7	b

If a particular Unicode hex value is not in this table, it may be that the word being analyzed is from another language that also uses Arabic Script for its writing, such as Farsi, Urdu, etc.

[α] Counting, of course, from the right, as any Arabic Script regardless of language is written from Right-to-Left.
[β] LibreOffice Writer or MS Word; in Softmaker's TextMaker, use Ctrl+Alt+Shift+X

چیدمان صفحه کلید فارسی

Farsi (Iranian/Persian) Layout

Pronunciation: cheedamon chataykedee farsee

← **IME Keystrokes**:]dnlhk wtpi ;gdn thvsd

Similar Glyphs / Letter Forms in the RTL Farsi/Persian Abjad N.B. Numeric Characters in Right-to-Left Languages are laid out in Left-to-Right order.

ى ی ي ئ ي	ث ت ب	خ ج ح	ض ص ش س	ظ ط	غ ع	ف ق ك ل ن	ة ه	لأ لآ لا
0649 0620 0626 064a	0628 062a 062b	062d 062c 062e	0633 0634 0635 0636	0637 0638	0639 063a	0641 0642 0643 0644 0646	0647 0629	fefb fef5 fef7
d S	f j e	p l o	s a w q	x z	u y	t r Z g k	i J	

Farsi is an Indo-European language used in Iran, Afghanistan (where it is called Dari), Tajikistan (where Tajiki is written with Cyrillic rather than Arabic script), and several other countries. There about 70 million speakers for whom Farsi is their primary or only language, and another 50 million or so speakers for whom Farsi is a second language. The Arabic ى (alef maksura u+649) looks identical to the Farsi ی (yeh u+6cc d) but has different initial and medial forms. The ة on the J key is a pre-composed glyph, and should ideally be entered as the components j N for future compatibility.

Note the paired delimiter reversals on the [,], O, P, {, and } keys; Farsi uses French style Guillemets (« and », located on the K and L keys) in place of parentheses. What look like stylized commas located on the English c, v and , keys are actually the alphabetic characters zain, reh and waw respectively. Also note that the question mark is reversed. shift+_ inserts a tatweel character (u+640), used as a "pseudo-kashida" prior to use of modern rendering engines.

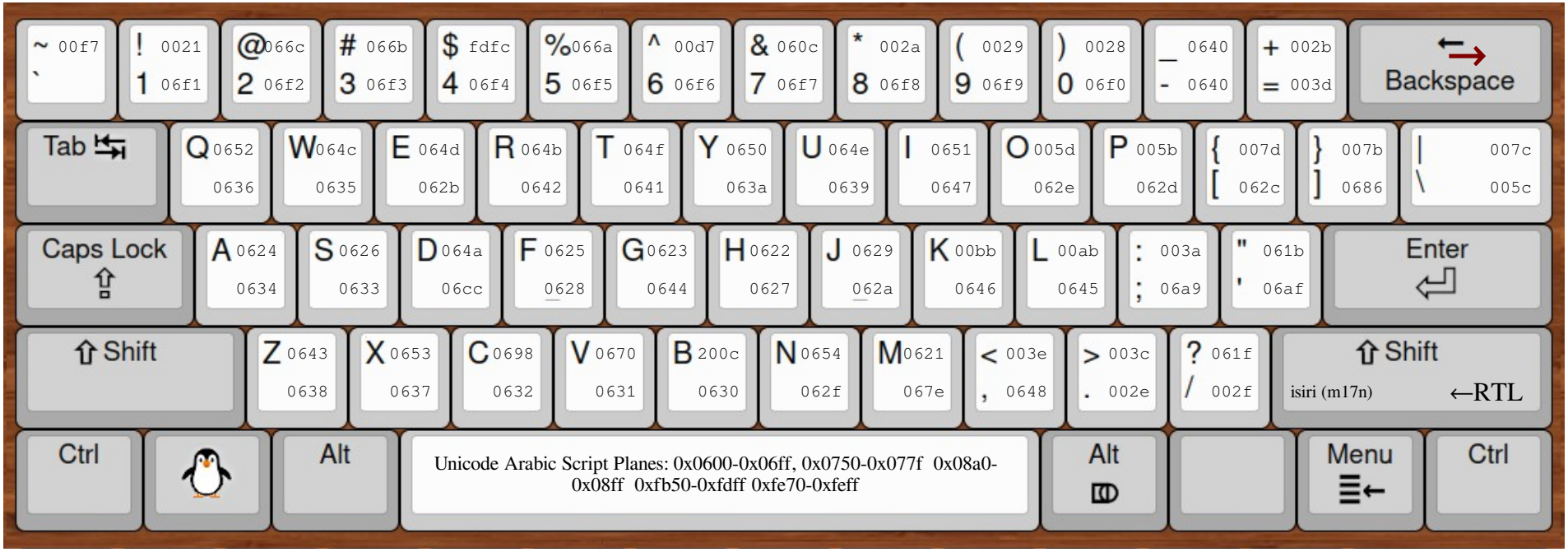
ZWJ on the B key is a Zero-Width Joiner; this non visible character forces adjacent characters to be joined that would otherwise not normally do so. The ZWNJ, shift+␣ on some IMEs, but not present on this keyboard layout, is a non-visible Zero-Width Non-Joiner character (u+200c) that does the opposite.

چیدمان صفحه کلید فارسی

Farsi (Iranian/Persian) Code Points

Pronunciation: cheedamon chataykedee farsee

← ***IME Keystrokes***:]dnlhk wtpi ;gdn thvsd



Farsi (Iranian / Persian) Text Analysis Example

To determine the key presses required to type کلید (key) – the third word[α] of the title – copy it into a word processor and move the cursor (▌) back and forth using the arrow keys, noting how, in RTL text, ← is interpreted as **forward** rather than **left**, etc.

e.g. ▌کلید ↔ د▌کلی ↔ ید▌کل ↔ لید▌ک ↔ کلید▌

Note that there are 4 characters. Now place the cursor after (*to the left of*) the first character. Use the Word Processor command Alt+x[β], to view the character's value.

The text will change from "لید▌ک" to "U+069a▌لید"; the initially disconcerting placement of the Unicode value of "ک" from the far right of the word to its left is because the string "U+069a" consists of all Latin LTR characters.

Without repositioning the cursor, press the forward/right arrow to go past each remaining character in sequence, pressing Alt+x[β] after each – the end result will be:

U+06a9U+0644U+06ccU+062f▌ – with all codes in LTR sequence.

Look up each Unicode value in the table's K column: u+6a9 is a semicolon, u+644 a g, and so forth; therefore کلید can be typed using the ;gdn key sequence.

[α] Counting, of course, from the right, since Farsi is written from Right-to-Left.
[β] LibreOffice Writer or MS Word; in Softmaker's TextMaker, use Ctrl+Alt+Shift+X

Unicode	UTF-8	K	Unicode	UTF-8	K	Unicode	UTF-8	K	Unicode	UTF-8	K	Unicode	UTF-8	K
u+60c	D88C	&	u+62c	D8AC	[	u+63a	D8BA	y	u+64e	D98E	U	u+6a9	DAA9	;
u+61b	D89B	"	u+62d	D8AD	p	u+640	D8C0	_	u+64f	D98F	T	u+6af	DAAF	'
u+61f	D89F	?	u+62e	D8AE	o	u+641	D8C1	t	u+650	D990	Y	u+6cc	DB8C	d
u+621	D8A1	M	u+62f	D8AF	n	u+642	D8C2	r	u+651	D991	I	u+6f0	DBB0	0
u+622	D8A2	H	u+630	D8B0	b	u+643	D8C3	Z	u+652	D991	Q	u+6f1	DBB1	1
u+623	D8A3	G	u+631	D8B1	v	u+644	D8C4	g	u+653	D992	X	u+6f2	DBB2	2
u+624	D8A4	A	u+632	D8B2	c	u+645	D8C5	l	u+654	D993	N	u+6f3	DBB3	3
u+625	D8A5	F	u+633	D8B3	s	u+646	D8C6	k	u+66a	D9AA	%	u+6f4	DBB4	4
u+626	D8A6	S	u+634	D8B4	a	u+647	D8C7	i	u+66b	D9AB	#	u+6f5	DBB5	5
u+627	D8A7	h	u+635	D8B5	w	u+648	D8C8	,	u+66c	D9AC	@	u+6f6	DBB6	6
u+628	D8A8	f	u+636	D8B6	q	u+64a	D98A	D	u+670	D9B0	V	u+6f7	DBB7	7
u+629	D8A9	J	u+637	D8B7	x	u+64b	D98B	R	u+67e	D9BE	m	u+6f8	DBB8	8
u+62a	D8AA	j	u+638	D8B8	z	u+64c	D98C	W	u+686	DA86	]	u+6f9	DBB9	9
u+62b	D8AB	e	u+639	D8B9	u	u+64d	D98D	E	u+698	DA98	C			

If a particular Unicode hex value is not in this table, e.g. u+649, it may be that the text being analyzed is from another language that also uses Arabic Script for its writing, such as Standard Arabic, Urdu, etc.

Ελληνικό πληκτρολόγιο

Pronunciation: Ellayniko playktrologio

Greek (Modern) Layout

← **IME Keystrokes**: Ellhnik;o plhktrol;ogio

The [:] key (¨) is the prefixed diacritic "combining diaresis." The [;] key (΄) is the prefixed diacritic "acute accent." Thus, to type ά, the [;] must be typed before the [a] key.

Note that, unlike modern Greek shown on this page, the IME keyboard for ancient/classical Greek (on page 11) requires that any diacritics be typed *after* the base character, i.e. the diacritics in ancient Greek are post-fixed.

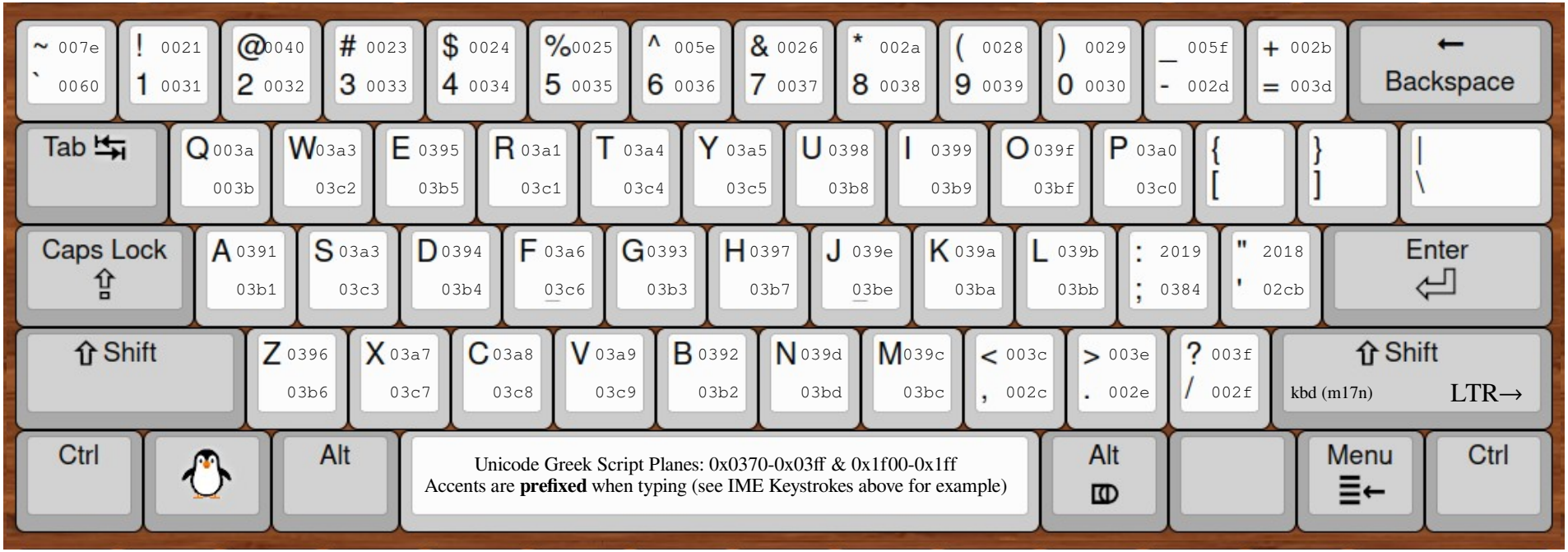
Unlike Arabic Scripts, the Greek alphabet has only one character with a 'final' form that is different from the norm. When the small sigma σ (typed with the [s] key) appears at the end of a word, it is written as an ς (on the [w] key). Unlike many IME engines, this is not handled automatically while typing, and the σ and ς appear as two distinct code points on disk.

Ελληνικό πληκτρολόγιο

Pronunciation: Ellayniko playktrologio

Greek (Modern) Code Points

← **IME Keystrokes**: Ellhnik;o plhktrol;ogio



As with most pre-computer era typewriter keyboards sold outside "the west," there are no Latin/English characters on the typewriter. Because the small (lower case) "l" key also served as the number 1 key on those typewriters (and the corresponding Greek lambda character "λ" was certainly inappropriate for this purpose), a "real" digit for one was added to Greek typewriters separately on the small/unshifted [9] key, as can be seen in the illustration on the right.

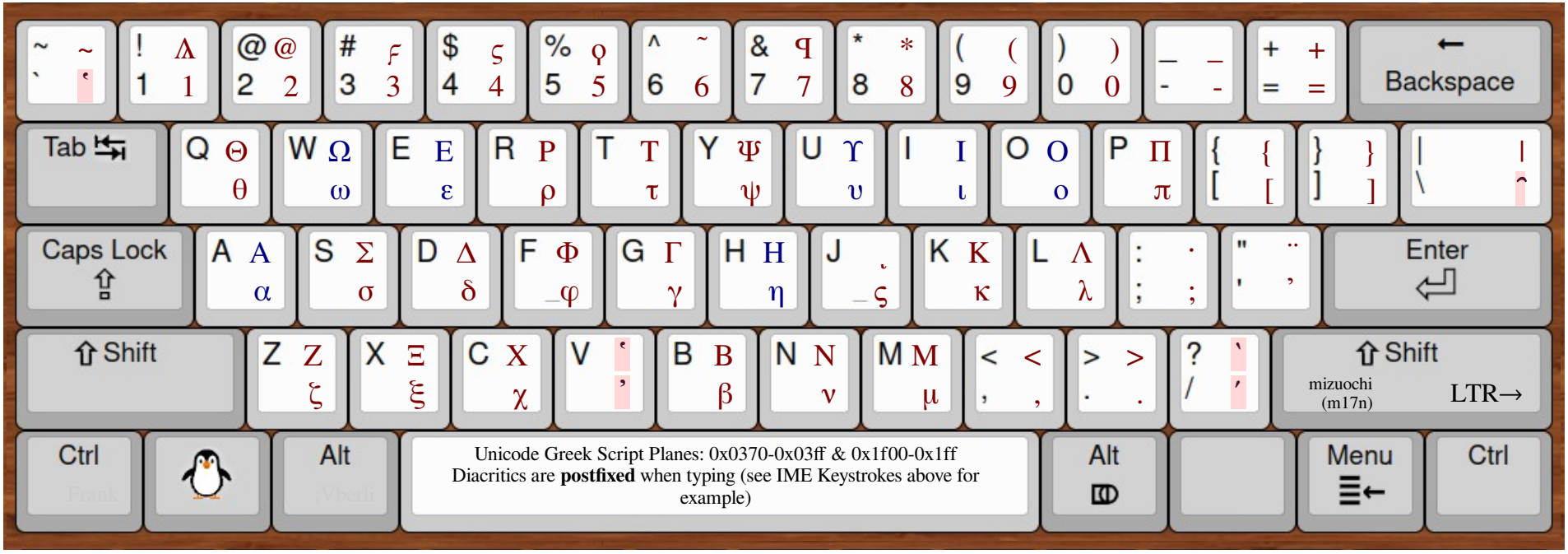
The slightly stylized Δρ produced by the capital [O] key is the formal abbreviation for **Δόκτορας** (a medical Doctor, similar to the English honorific "Dr."); this same symbol is also used in the scientific community for normalized magneto-resistance. I've been unable to determine why either of these merited a place on a popular production typewriter though.

Ελληνικό πληκτρολόγιο

Pronunciation: Ellayniko playktrologio

Ancient Greek (Classical – Mizuochi) Layout

← **IME Keystrokes**: Ellhnik;o plhktrol;ogio



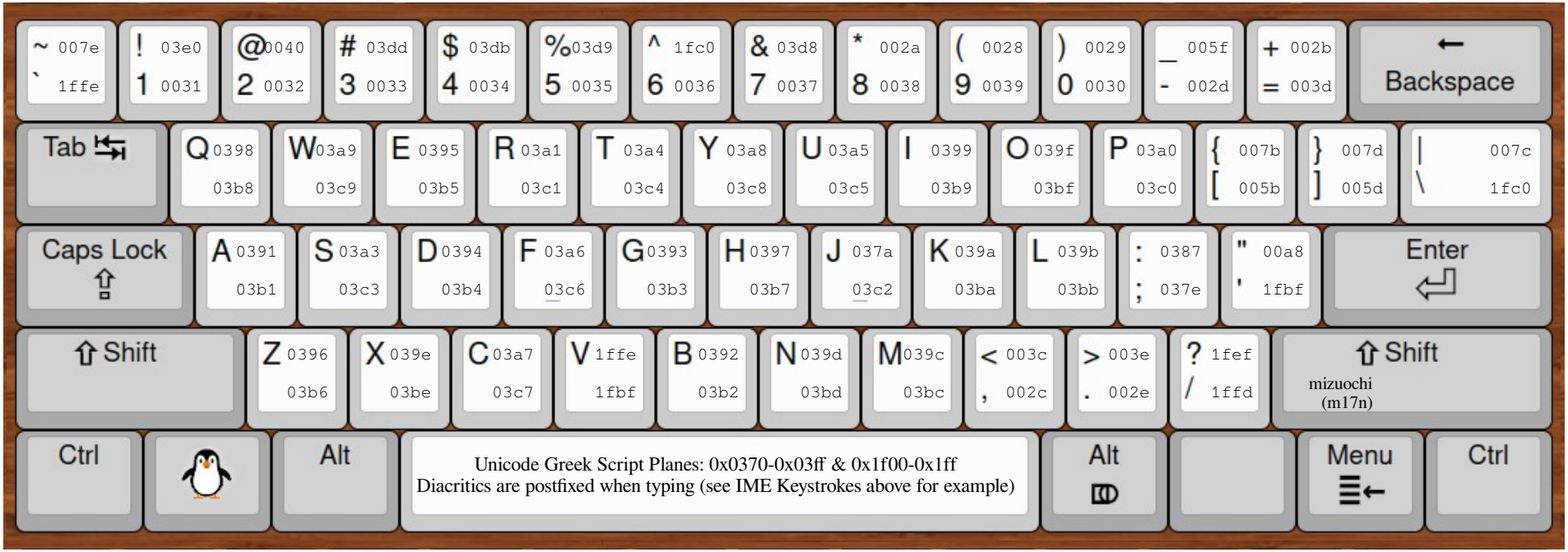
Ancient Greek Alphabet Typing Post-Fix Accents & Breathing Marks		α a 03b1	β b 03b2	γ g 03b3	δ d 03b4	ε e 03b5	ζ z 03b6	η h 03b7	θ u 03b8	ι i 03b9	κ k 03ba	λ l 03bb	μ m 03bc	ξ x 03be	ν n 03bd	ο o 03bf	π p 03c0	ρ r 03c1	σ s 03c3	ς j 03c2	τ t 03c4	υ u 03c5	φ f 03c6	χ c 03c7	ψ y 03c8	ω w 03c9
Acute Accent *όξεα*	+ /	ά 1f71				έ 1f73		ή 1f75		ί 1f77						ό 1f79						ύ 1f7b				ώ 1f7d
Grave Accent *υαρεῖα*	+ ?	ὰ 1f70				ὲ 1f72		ὴ 1f74		ὶ 1f76	*The grave accent may appear only over the last syllable as a replacement for an acute.*					ὸ 1f78						ὺ 1f7a				ὼ 1f7c
Circumflex Accent *περισπομένη*	+ \	ᾶ 1fb6				ῆ 03b5 0342		ῆ 1fc6		ῖ 1fd6	*Don't confuse the tilde ~ (upper case `) with the modern circumflex ~ (upper case 6)*					ὸ 03bf 0342	*Only with the ε and ο are the actual entries retained on disc and dynamically rendered.*					ῦ 1fe6	*ᾶ ῆ ῖ ῆ ῦ & ῶ are replaced on disk with composites*		ῶ 1ff6
Smooth Breathing *ψιλί*	+ v + '	ἀ 1f00	*Aspirations are on the 2nd vowel of dipthongs.*			ἐ 1f10		ἠ 1f20		ἰ 1f30						ὀ 1f40						ὐ 1f50				ὠ 1f60
Rough Breathing *δασεῖα*	+ V + `	ἁ 1f01	*Aspirations are on the 2nd vowel of dipthongs.*			ἑ 1f11		ἡ 1f21		ἱ 1f31						ὁ 1f41						ὑ 1f51	*See ^ accent with aspirations on the next page.*			ὡ 1f61

Ελληνικό πληκτρολόγιο

Pronunciation: Ellayniko playktrologio

Ancient Greek (Mizuochi) Code Points

← **IME Keystrokes**: Ellhnik;o plhktrol;ogio



Reality Check: Greek Diacritics, such as accents and breathing marks only began to be widely used in the 3rd century BCE, (i.e. 300bce-201bce).

The poet Homer probably lived between 1200 and 701bce. Xenophon lived from 430bce to 354bce, and Aristotle from 384bce through 322bce.

Homer – being blind – didn't actually write anything himself, and what we have from Aristotle might possibly be class notes from one of his students.

So: Diacritics (or spaces between words, or lower case letters) would not have been used by these authors. It is traditional, however, to use these elements in modern renderings of "Classic" authors, including the multitude of other Hellenic playwrights, mathematicians and philosophers.

Combining Perispomeni with Aspirations

The circumflex or perispomeni is the only accent that ever appears with breathing marks, and is always entered before the aspiration symbol.

a \ v	a \ V	A \ v	A \ V	h \ v	h \ V	H \ v	H \ V	i \ v	i \ V
ἆ	ἇ	Ἆ	Ἇ	ἦ	ἧ	Ἦ	Ἧ	ἶ	ἷ
1f06	1f07	1f0e	1f0f	1f26	1f27	1f2e	1f2f	1f36	1f37

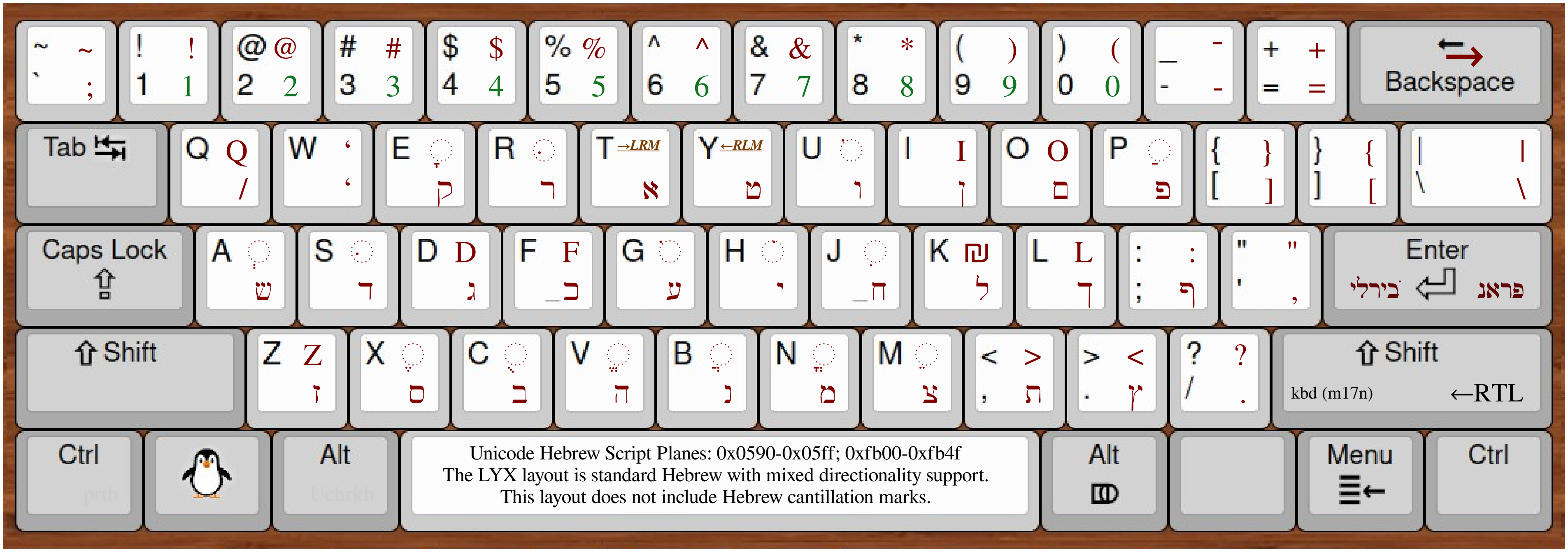
I \ v	I \ V	u \ v	u \ V	U \ V	w \ v	w \ V	W \ v	W \ V	
Ἶ	Ἷ	ὖ	ὗ	Ὗ	ὦ	ὧ	Ὦ	Ὧ	
1f3e	1f3f	1f56	1f57	1f5f	1f66	1f67	1f6e	1f6f	

פריסת מקלדת עברית

Pronunciation: porisak neeklebit eevreek

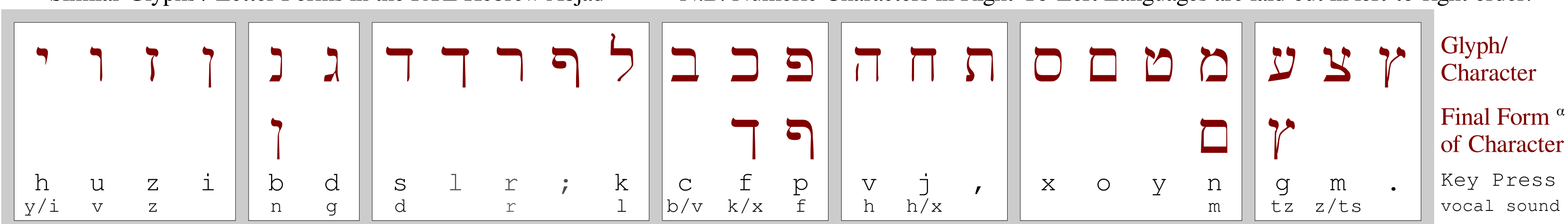
Hebrew (m17n) Layout

← ***IME Keystrokes***: prhx, neks, gcrh,



Similar Glyphs / Letter Forms in the RTL Hebrew Abjad

N.B. Numeric Characters in Right-To-Left Languages are laid out in left-to-right order.



Note the paired delimiter reversals on the [(], [)], [[], []], [{], and [}] keys; Hebrew treats the keys correctly as “open/close” rather than “right/left” characters.

Do not confuse the Maqaf punctuation character ־ (on the shifted [_] key) with the similar hyphen - (on the unshifted [-] key)

Unicode Left-to-Right Marker [T] (u+200e) and Right-to-Left Marker [Y] (u+200f), useful for short lengths of text, are both invisible; these help, but cursor motion between words is still decidedly non-intuitive because, in my opinion, spaces in particular should not be treated as if the typist were “returning to Latin” but as neutral, non-specific characters as prescribed in the Unicode Bidirectional Algorithm (Unicode® Standard Annex #9, https://unicode.org/reports/tr9/.)

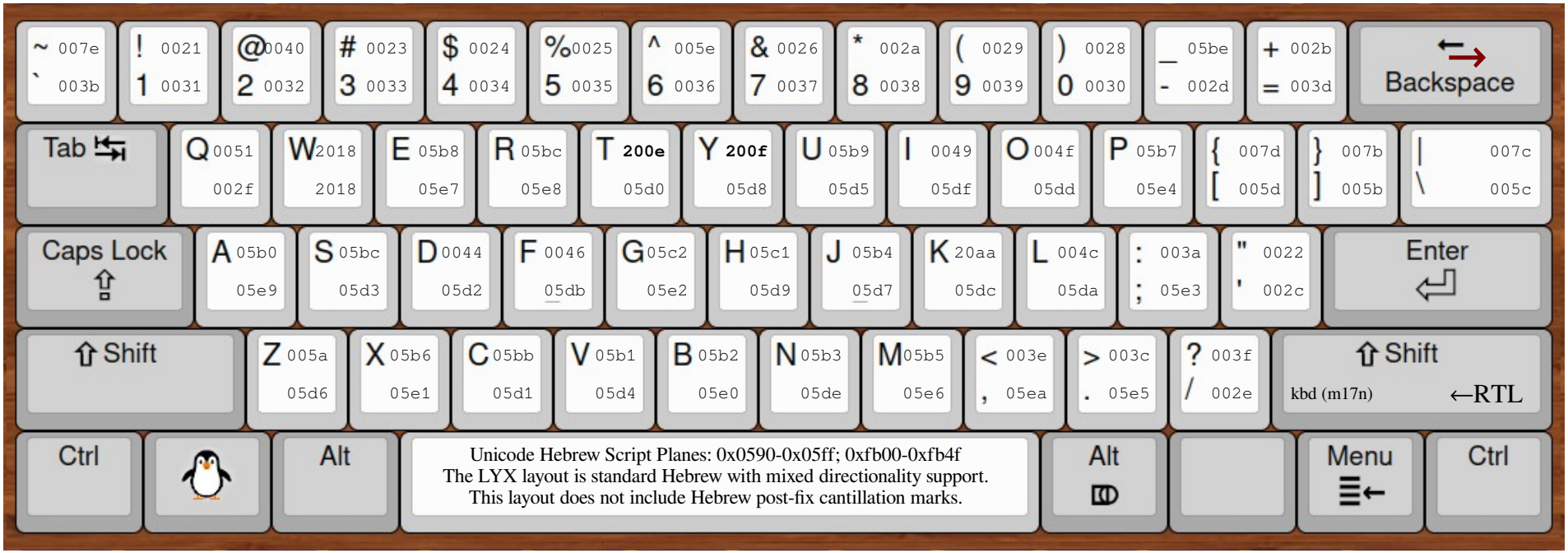
α Two characters in Hebrew have “final forms” (כ and פ), but thse are applied automatically by most Input Method Editors and shaping engines.

פרימת מקלדת עברית

Pronunciation: porisak neeklebit eevreek

Hebrew (m17n) Code Points

← **IME Keystrokes**: prhx, neks, gcrh,



This Swiss-made Hebrew typewriter has the Israeli Lira £ and U.S. Dollar $ symbols on the top left key, with the Israeli Pruta ל"י mark on the key below. The אג. on the shifted portion of the latter key is an abbreviation for a formal greeting akin to "Dear" in English – usually used with אדון or גברת (Sir or Madam).

The primary alphabetic letters, however, remained in the same positions when computer keyboards began to appear.

The small (lower case) "L" key also serves as the number 1 key on western typewriters but since there is no appropriate Hebrew character, a "real" digit "1" was added to Hebrew typewriters as illustrated on the right.

Typing order where diacritics are supported is letter first, then optional dagesh (indicating a hard or soft consonant), then any additional marks, such as cantillation symbols, etc.

See DDN-5 (see page 2) or https://dotancohen.com/howto/rtl_right_to_left.html for a good discussion of bidirectionality issues.

हिंदी कीबोर्ड लेआउट

Pronunciation: hindi keebord leaut

Hindi (Inscript-Standard) Layout

← **IME Keystrokes**: ufxor krya[nsEG'

Similar Letter Forms in Devanagari Script:

The Inscript keyboard layout utilizes several keys as "macro" keys, such as "V" for "v+]", "A" for "E+s", etc. The name

फ़ऐक ओबरली

is not typed with the individual keys:

H]Ssk Esyjnr

but by substituting the macro keys W and A:

H]Wk Ayjnr

ट '	ठ "	ग्र u97a	व b	च ;	ष <	ब y	ब्र u97f	क k	क़ k]		ञ }	ऋ u90b	ॠ u960		। >	॥ u965	र j	ऱ **J** j]		ख K	ख़ K]	*Cells with Red borders are "macro" keys.*
ह u	ढ {	ढ़ {]	द o		न v	ऩ **V** v]	न् u97b	त l	ग i	ज p	ज़ p]	ऌ u979	ॡ u97c		ळ N	ऴ B	क़ N]		उ G	ऊ T	ल n	ॢ u90c	ॣ u961	
अ D	अं Dx	अः D_	अ D@	अ Dz		आ E	ऑ | E@	ओ ~	आ Ez	औ u975	ओ **A** Es	औ **Q** Ew	आ u974	आं Ex		ध O	घ I	घ़ I]	छ :	क्ष **&** kd<	*Not all characters in this table are used in the Hindi alphabet and/or Hindi keyboard.*	
भ Y	म c	य /	य़ /]	स m	थ L	प h	फ H	फ़ H]	ण C		ऐ S	ऍ ! S@	ऐ **Z** Sz	ऐ **W** Ss		इ F	ई R	ङ U	झ P	ड [	ड़ []	ऽ u93d	ङ़ u97e	

There are more than 258 million speakers in India whose primary language is Hindi and another 214 million L2 speakers for whom it is a second language.

The Ctrl+Shift+I sequence on many Inscript keyboards generates a ZWJ (u+200d), although the I is sometimes used. See page 7 for information on ZWJ.

हिंदी कीबोर्ड लेआउट

Pronunciation: hindi keebord leaut

Hindi (Inscript-Standard) Code Points

← IME Keystrokes: ufxor krya[nsEG'

Hindi Text Analysis Example

To determine the key presses required to type हिंदी – the first word of the title above – copy it into a word processor and move the cursor (▌) back and forth using the arrow keys:

e.g. ▌हिंदी ↔ हिं▌दी ↔ हिंदी▌

There *seem to be* 2 characters, but place the cursor after the first character and type Alt+x[α], to view the "character's" value. The text changes from " हिं | दी " to "U+0939U+093fU+0902 | दी ", indicating that हिं is a composite character – akin to a ligature.

By looking up those values in the table to the right, we learn that the sequence ufx was used to type ह, ि, and ं in that order. Not only have the characters been combined, but ि has been swapped with ह. Remember, though, that the original three character sequence is what is stored on disk.

Repeating the process with the next character दी ▌ indicates that these 2 characters are typed with o and r.

If a particular code is not in this table, the word may be from another language that uses Devanagari Script to write its text, such as Konkani, Marathi, Nepali, etc.

Unicode	UTF-8	K
u+901	A981	X
u+902	A482	x
u+903	A483	_
u+905	A485	D
u+906	A486	E
u+907	A487	F
u+908	A488	R
u+909	A489	G
u+90a	A48A	T
u+90b	A48B	+
u+90d	A48D	!
u+90e	A48E	Z
u+90f	A48F	S
u+910	A490	W

Unicode	UTF-8	K
u+911	A491	|
u+912	A492	~
u+913	A493	A
u+914	A494	Q
u+915	A495	k
u+916	A496	K
u+917	A497	i
u+918	A498	I
u+919	A499	U
u+91a	A49A	;
u+91b	A49B	:
u+91c	A49C	p
u+91d	A49D	P
u+91e	A49E	}

Unicode	UTF-8	K
u+91f	A49F	'
u+920	A4A0	"
u+921	A4A1	[
u+922	A4A2	{
u+923	A4A3	C
u+924	A4A4	l
u+925	A4A5	L
u+926	A4A6	o
u+927	A4A7	O
u+928	A4A8	v
u+929	A4A9	V
u+92a	A4AA	h
u+92b	A4AB	H
u+92c	A4AC	y

Unicode	UTF-8	K
u+92d	A4AD	Y
u+92e	A4AE	c
u+92f	A4AF	/
u+930	A4B0	j
u+931	A4B1	J
u+932	A4B2	n
u+933	A4B3	N
u+934	A4B4	B
u+935	A4B5	b
u+936	A4B6	M
u+937	A4B7	<
u+938	A4B8	m
u+939	A4B9	u
u+93c	A4BC	]

Unicode	UTF-8	K
u+93e	A4BE	e
u+93f	A4BF	f
u+940	A580	r
u+941	A581	g
u+942	A582	t
u+943	A583	=
u+945	A585	@
u+946	A586	z
u+947	A587	s
u+948	A588	w
u+949	A589	\
u+94a	A58A	`
u+94b	A58B	a
u+94c	A58C	q

Unicode	UTF-8	K
u+94d	A58D	d
u+95f	A59F	?
u+964	A5A4	>
u+966	A5A6	0
u+967	A5A7	1
u+968	A5A8	2
u+969	A5A9	3
u+96a	A5AA	4
u+96b	A5AB	5
u+96c	A5AC	6
u+96d	A5AD	7
u+96e	A5AE	8
u+96f	A5AF	9

[α] LibreOffice Writer or Microsoft Word; in Softmaker's TextMaker, use Ctrl+Alt+Shift+X

ひらがなキーボード

Pronunciation: Hiragana kībōdo

Japanese (Romaji→Hiragana) Layout

← IME Keystrokes: vot[u *Enter* g_-[_s[*F7 Enter*

Entering Hiragana (shown above) and/or Katakana with the Romaji Typing Method

Unlike most languages, Japanese can be (and often is) written using multiple Scripts – Hiragana, Katakana, Kanji, and Romaji – often in a single sentence. Regardless, Japanese is most easily entered on modern devices using the phonetic Romaji layout, with the IME handling conversion to the required output. For this technique, use the following Japanese IME settings:

keyboard=English (U.S.), input mode=Hiragana, typing method=Romaji

To type the seven syllable title "hiragana kībōdo" (ひらがなキーボード), enter its Romaji spelling as if it were English using the following keystrokes:

h i r a g a n a k i - b o - d o Enter

The macron over the ī and ō, which indicate that these vowels are to be elongated are shown in Hiragana with the post-fix ー character, which is entered using the - key.

The result is ひらがなきー ぼー ど; although the first four syllabic characters match, three of the remaining characters (き, ぼ, and ど) don't resemble the desired キ, ボ, and ド.

The reason is that, since "keyboard" is an adopted foreign word, it would commonly be written with Katakana rather than Hiragana script. With many IMEs, a "non-commited" stretch of Hiragana can be converted to Katakana by using the F7 key. The solution therefore is to press Enter after typing the first word to "commit" it, then pressing F7 after the second to convert it to Katakana, followed by Enter to commit it. This is illustrated in the following example:

h i r a g a n a Enter k i - b o - d o F7 Enter

This results in the correct display ひらがなキーボード, a not atypical mixture of two different writing systems even in a single Japanese phrase.

Useful command keys available during Romaji entry (and prior to any "commit") include:

- F6 - convert an uncommited Katakana phrase to its Hiragana representation
- F7 - convert an uncommited Hiragana phrase to its Katakana representation
- F8 - convert an uncommited Hiragana phrase to its half-width character versions
- Space - convert an uncommited Hiragana or Katakana phrase to its Kanij representation. Depending on IME settings a menu of possible conversion options will be shown.

Note that other IMEs may use different commands than these iBus-Anthy examples.

Kībōdo (キーボード), the word used for typewriter and computer keyboards, is a good example of the primary use for Katakana. While this could be converted to Kanji (鍵盤, pronounced kenban) by pressing Space before "committing," that would be inappropriate, since 鍵盤 refers to more traditional "keyboards," such as those on musical instruments.

There are multiple number schemes used in Japanese as well: The Western (Arabic) numerals, the original Sino-Japanese, and the native kun'yomi; the arabic numerals are generally used in horizontal texts, while the Kanji numerals are often seen in vertical texts. Dot not confuse the Kanji 一 (the number one) with the elongation symbol ー on the - key.

ひらがなキーボード

Pronunciation: Hiragana kībōdo

Japanese (Kana) Layout

← IME Keystrokes: vot[u *Enter* g_-[_s[*F7 Enter*

Entering Hiragana and/or Katakana with the Kana Typing Method (Hiragana keys shown above)

Each Hiragana or Katakana glyph (collectively known as "kana") represents a single syllable or diacritic, and the more direct Kana typing method assigns most Japanese syllables to a single key or key pair rather than relying on the Romaji method's "translation." For this technique, use the following Japanese IME settings:

keyboard=English (U.S.), input mode=Hiragana, typing method=Kana

To begin typing the title **ひらがなキーボード** shown above, either locate the ひ on the keyboard layout or find its phonetic representation HI in the table on the right; in either case the ひ syllable is typed using the lower case [v] key. Similarly, the second syllable ら (RA) is typed with the [o] key. The third syllable が (GA), however, cannot be found on the keyboard; か is a voiced version of か (KA), created by following it with the "voicing" diacritic[α] ゛ located on the [[] key. This is why certain sounds are shown with multiple key strokes in the table on the right. The final syllable is of course entered with [u].

Because kībōdo is an adopted word – which are traditionally written using Katakana – the [Enter] will "commit" ひらがな as Hiragana so that きーぼーど can be typed separately, converted to キーボード with the [F7] key and committed as a Katakana spelling. kī is typed with the [g] key followed by the vowel lengthening sign ー on the [-] key. The syllable bo (ぼ, a voiced ho/ほ, and therefore entered with the [-][[] sequence is also followed by a vowel lengthening mark ー. After the final do (ど, a voiced to/と), [F7] performs the conversion to Katakana, and [Enter] commits the word. The complete sequence is:

[v][o][t][[][u][Enter][g][_][-][[][_][s][[][F7][Enter]

See other comments on mixing Hiragana and Katakana on page 17.

[α] The ゛ voicing diacritic is known as a Dakuten; the ゜ on the [[] key is called a Handakuten.

Phoneme	K	Phoneme	K	Phoneme	K	Phoneme	K	Phoneme	K
A	3	GU ぐ	h[	LTSU	z	NI	i	TA	q
BI び	v[	HA	f	LU	$	NN	y	TE	w
BO ぼ	-[	HE	=	LYA	&	NO	k	TO	s
BU	<[	HI	v	LYE	%	NU	1	TSU	z
CHI	a	HO	-	LYI	E	O	6	U	4
DA	q]	HU	2	LTU	Z	PI	v]	WA	0
DE で	w[	I	e	LYO	(	RA	o	WO	)
DI ぢ	a[	JI じ	d[	LYU	*	RE	;	WYE	+
DO ど	s[	KA	t	MA	j	RI (kata)	l	YA	7
DU づ	z[	KE	'	ME	/	RO	`	YO	9
E	5	KI	g	MI	n	RU	.	YU	8
FU (kata)	2	KO	b	MO	m	SA	x	ZA	r]
GA が	t[	KU	h	MU	\	SE	p	ZE ぜ	p[
GE げ	'[	LE	%	NA	u	SHI	d	ZO ぞ	c[
GI ぎ	g[	LI	E	NE	,	SO	c	ZU ず	r[
GO ご	b[	LO	^	NGA	f[	SU	r	ー	_

ひらがなキーボード

Japanese (Hiragana) Code Points

Pronunciation: Hiragana kībōdo

← IME Keystrokes: vot[u *Enter* g_-[_s[*F7 Enter*

Japanese Hiragana Text Analysis Example

To determine the key presses required to type ひらがな – the first word of the title above – copy it into a word processor and move the cursor (▌) back and forth using the arrow keys:

e.g. ▌ひらがな ↔ ひ▌らがな ↔ ひら▌がな etc.

There seem to be 4 characters. Place the cursor after each of the first two characters and type Alt+x[α] to view the characters' values. The text becomes "U+3072U+3089▌がな" – the table on the right indicates that ひ and ら are typed with the [v] and [o] keys respectively. Continuing this process results in:

U+3072U+3089U+304cU+306a▌

Repeating this process for キーボード, the remainder of the title, indicates that these values are all Katakana characters: not unexpected, since kībōdo (say it out loud) is a "borrowed" word. The analysis, however, can continue in the same manner by subtracting 0x60 from each hex value. See page 20.

Unicode	UTF-8	K
u3001	E38081	<
u3002	E38082	>
u300c	E3808C	{
u300d	E3808D	}
u3041	E38181	#
u3042	E38182	3
u3043	E38183	E
u3044	E38184	e
u3045	E38185	$
u3046	E38186	4
u3047	E38187	%
u3048	E38188	5
u3049	E38189	^
u304a	E3818A	6
u304b	E3818B	t

Unicode	UTF-8	K
u304c	E3818C	t[
u304d	E3818D	g
u304f	E3818F	h
u3051	E38191	'
u3052	E38192	'[
u3053	E38193	b
u3054	E38194	b[
u3055	E38195	x
u3057	E38197	d
u3058	E38198	d[
u3059	E38199	r
u305a	E3819A	r[
u305b	E3819B	p
u305c	E3819C	p[
u305d	E3819D	c

Unicode	UTF-8	K
u305e	E3819E	c[
u305f	E3819F	q
u3060	E381A0	q[
u3061	E381A1	a
u3062	E381A2	a[
u3063	E381A3	z
u3064	E381A4	z
u3066	E381A6	w
u3067	E381A7	w[
u3068	E381A8	s
u306a	E381AA	u
u306b	E381AB	i
u306c	E381AC	1
u306d	E381AD	,
u306e	E381AE	k

Unicode	UTF-8	K
u306f	E381AF	f
u3072	E381B2	v
u3075	E381B5	2
u3078	E381B8	=
u307b	E381BB	-
u307e	E381BE	j
u307f	E381BF	n
u3080	E381C0	\
u3081	E381C1	/
u3082	E381C2	m
u3083	E381C3	&
u3084	E381C4	7
u3085	E381C5	*
u3086	E381C6	8
u3087	E381C7	(

Unicode	UTF-8	K
u3088	E381C8	9
u3089	E381C9	o
u308a	E381CA	l
u308b	E381CB	.
u308c	E381CC	;
u308d	E381CD	`
u308f	E381CF	0
u3091	E381D1	+
u3092	E381D2	)
u3093	E381D3	y
u309b	E381DB	[
u309c	E381DC	]
u30fb	E383BB	?
u30fc	E383BC	-

[α] LibreOffice Writer or MS Word; in Softmaker's TextMaker, use Ctrl+Alt+Shift+X

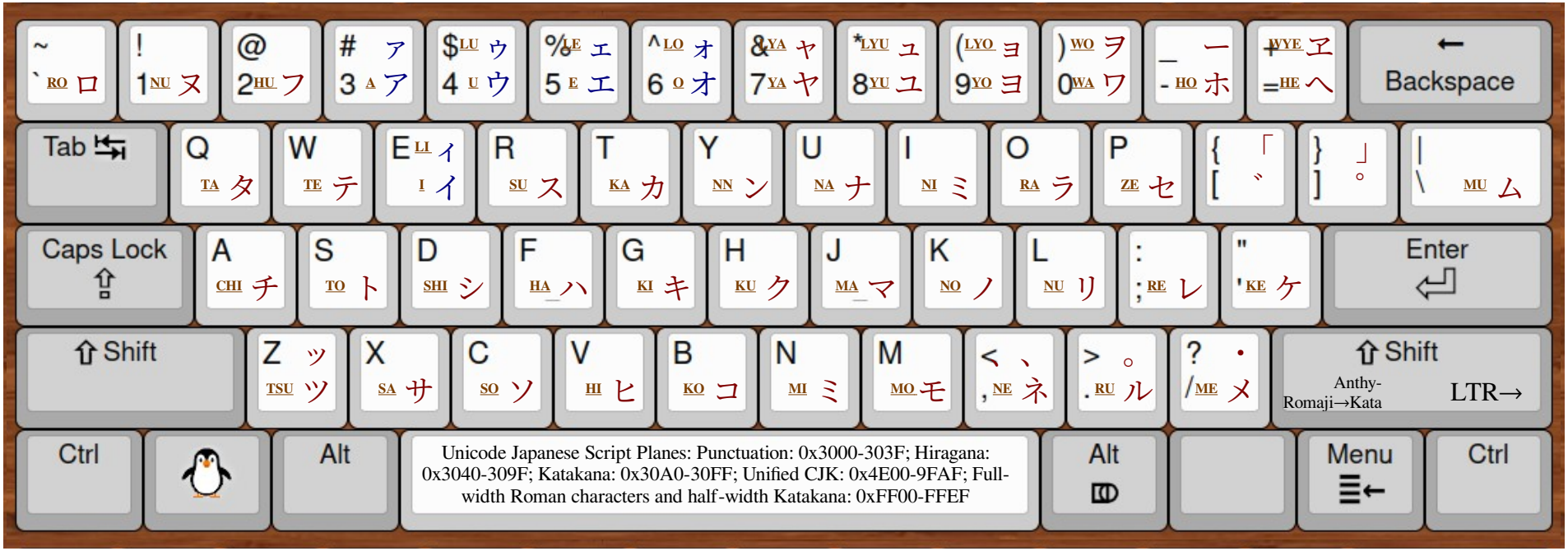
See https://www.coscom.co.jp/learnjapanese801/lesson08.html for converting romaji to kanji ...also see: https://japaneseup.com/typing-japanese-letters/

キーボードと言う

Japanese (Romaji→Katakana) Layout

Pronunciation: Kībōdo to iu (FreeSerif 22 v Noto 20)

← IME Keystrokes: g_-[_s[??$ (remainder is Katakana)



Entering Katakana (shown above) and/or Hiragana with the Romaji Typing Method

Because Katakana is used for words adopted from other languages, it is more common to use the Romaji-to-Hiragana keyboard layout presented on page 17 on modern computing devices. This layout page is included only to illustrate the correspondence between the Hiragana and Katakana syllabic characters on the keyboard. Though only relevant to direct Kana input layout (see page 18), the phonetic representations for each syllable are also included on this layout for comparison/convenience.

The only significant difference between the Hiragana and Katakana layouts to be aware of is the mapping of the hyphen/underscore key [-] on the top row (between the [0] and [=] keys). With that exception, all of the code points and corresponding key presses are the same as on page 19, but 0x60 higher (see the Hiragana/Katakana comparison table in DDN-7, "Exploring Han Script Entry – Chinese", for more detail). The Hiragana code points u+3041 through u+3093, for example, correspond to Katakana code points u+30a1 through u+30f3. The voiced sound mark u+309b, semi-voiced sound mark u+309c, elongated sound mark, u+30fb, and middle dot u+30fc are shared across both Hiragana and Katakana Scripts.

To use this technique for direct entry of Katakana syllabic characters, apply the following Japanese IME settings:

keyboard=English (U.S.), input mode=Katakana, typing method=Romaji

A code point page is not provided for Katakana Script key presses, since these can be calculated by adding 0x60 to each of the values shown for Hiragana Script on page 19.

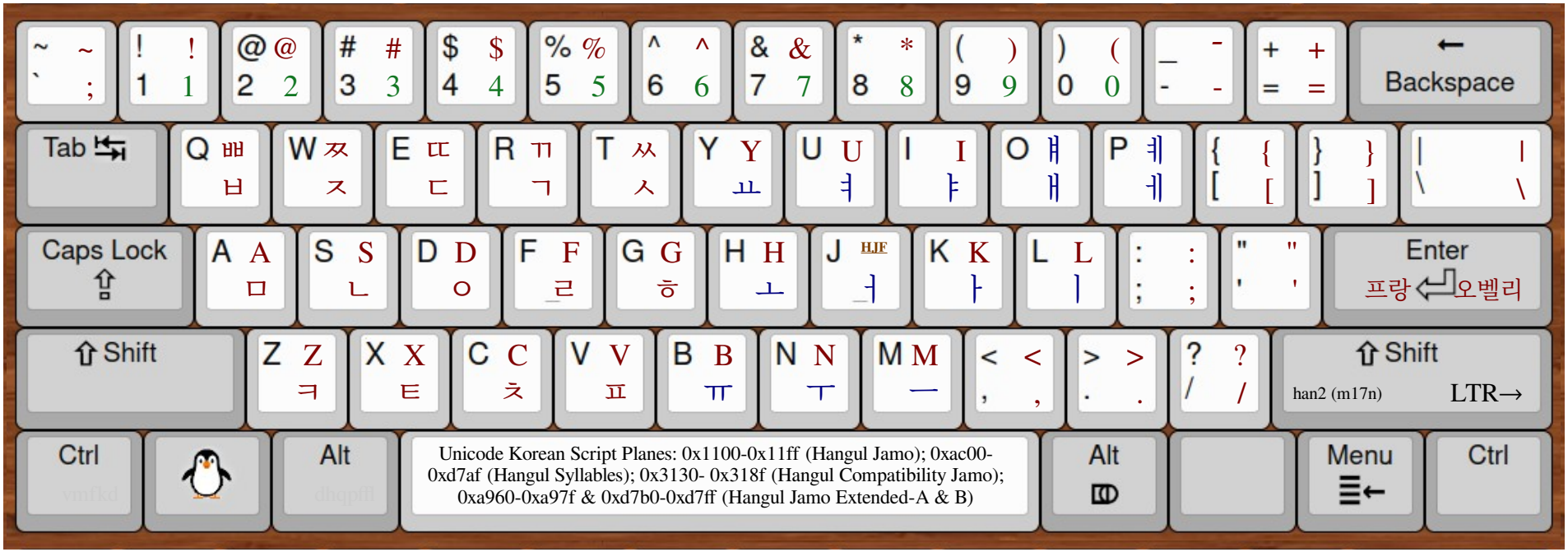
DDN-7 "Exploring Han Script Entry – Chinese," though primarily targeted toward Chinese writing, includes discussion of other Asian languages that adopted forms of Hànzì in the past, including Japanese – where it still survives in the form of Kanji.

한국어 키보드 레이아웃

Korean (Hangul-2 Jamo) Layout

Pronunciation: hangug-eo kibodeu leiaus

← ***IME Keystrokes***: gksrnrdj zlqhem fpdjdkdmt



Similar Jamo Glyphs / Component Letter Forms in the Korean Alphabet with key presses and pronunciation. **Consonants (on left) are shown in Red, Vowels (on right) in Blue.**

ㅇ	ㅎ	ㅅ	ㅆ	ㅊ	ㅈ	ㅉ	ㄴ	ㄷ	ㄸ	ㅁ	ㅌ	ㅋ	ㄱ	ㄲ	ㄹ	Glyph/Character
d	g	t	T	c	w	W	s	e	E	a	x	z	r	R	f	Key Press Used
king	hope	soul	sea	chick	chill	midget	nose	tall	steam	mother	talk	king	kiss	skill	lung	vocal sound

ㅖ	ㅒ	ㅕ	ㅓ	ㅣ	ㅏ	ㅑ	ㅐ	ㅔ	ㅡ	ㅗ	ㅛ	ㅠ	ㅜ	ㅂ	ㅃ	Glyph/Character
P	p	u	j	l	k	i	o	O	m	h	y	v	b	n	q	Q	Key Press Used
yes	met	yon'	uh	feet	ah	yard	care	yank	pull	oh	yoga	pill	you	boo	park	speak	vocal sound

Dubeolsik (두벌식), the most common Hangul keyboard layout has consonants on the left and vowels on the right.
HJF is the HANGUL JUNGSEONG FILLER (u+1160), a medial vowel, is blacklisted by most browser address parsing routines, though the reason(s) why this particular character is uniquely dangerous is unclear to me. It is related to the HF Hangul Filler (u+3164), but information about these characters is sparse!!

한국어 키보드 레이아웃

Korean (Hangul-2 Jamo) Code Points

Pronunciation: hangug-eo kibodeu leiaus

← IME Keystrokes: gksrnrdj zlqhem fpdjdkdmt

Order/Index	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Lead/Initial Consonant	ㄱ	ㄲ	ㄴ	ㄷ	ㄸ	ㄹ	ㅁ	ㅂ	ㅃ	ㅅ	ㅆ	ㅇ	ㅈ	ㅉ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ								
Key Press	r	R	s	e	E	f	a	q	Q	t	T	d	w	W	c	z	x	v	g								
Jamo U+11+	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	10	11	12								
(Medial) Vowel	ㅏ	ㅐ	ㅑ	ㅒ	ㅓ	ㅔ	ㅕ	ㅖ	ㅗ	ㅘ	ㅙ	ㅚ	ㅛ	ㅜ	ㅝ	ㅞ	ㅟ	ㅠ	ㅡ	ㅢ	ㅣ						
Key Press(es)	k	o	i	O	j	p	u	P	h	hk	ho	hl	y	n	nj	np	nl	b	m	ml	l						
Jamo U+11+	61	62	63	64	65	66	67	68	69	6A	6B	6C	6D	6E	6F	70	71	72	73	74	75						
Ending/Tail Consonant	ㄱ	ㄲ	ㄳ	ㄴ	ㄵ	ㄶ	ㄷ	ㄹ	ㄺ	ㄻ	ㄼ	ㄽ	ㄾ	ㄿ	ㅀ	ㅁ	ㅂ	ㅄ	ㅅ	ㅆ	ㅇ	ㅈ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ
Key Press(es)	r	R	rt	s	sw	sg	e	f	fr	fa	fq	ft	fx	fv	fg	a	q	qt	t	T	d	w	W	z	x	v	g
Jamo U+11+	A8	A9	AA	AB	AC	AD	AE	AF	B0	B1	B2	B3	B4	B5	B6	B7	B8	B9	BA	BB	BC	BD	BE	BF	C0	C1	C2

The language of 𐀔𐀪𐀞 (Knossos)

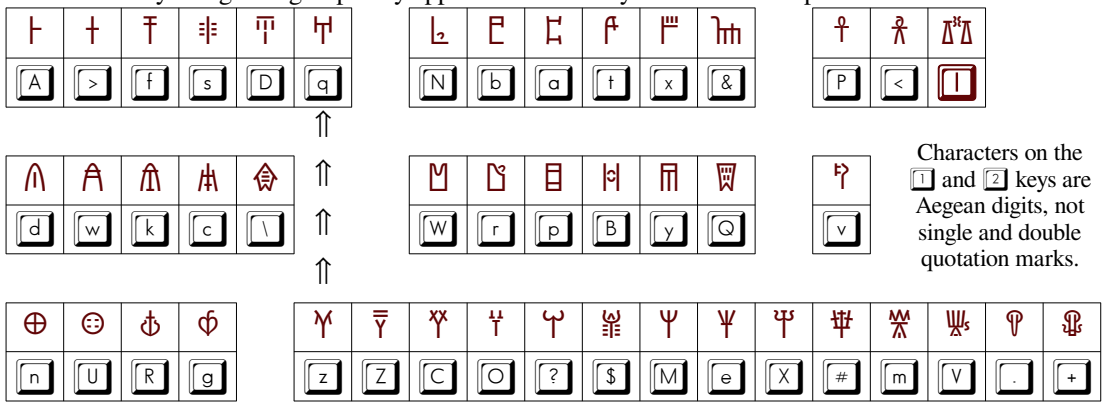
Mycenean (Linear B Syllabary) Layout

Pronunciation: Kaw Naw Saw (the capitol of ancient Crete)

← IME Keystrokes: .Vv

Keyboard for Linear-B Syllabary; extension of Linear B Keyboard Layout © 2014, Richard Vallance Janke, for use with my extended method of numeric entry.

Linear B Syllabograms grouped by appearance with Key Presses used to produce them



Characters on the 1 and 2 keys are Aegean digits, not single and double quotation marks.

Linear B Symbols with Bennett Codes

u+10050		B018	u+10057		B063
u+10051		B019	u+10058		B064
u+10052		B022	u+10059		B079
u+10053		B034	u+1005A		B082
u+10054		B047	u+1005B		B083
u+10055		B049	u+1005C		B086
u+10056		B056	u+1005D		B089

alt+ produces u+10100, the word separator line; Ctrl+ produces u+10101 (·), the word separator dot.

The language of 𐀏𐀺𐀍 (Knossos)

Mycenean (Linear B Syllabary) Code Points

Pronunciation: Kaw Naw Saw (the capitol of ancient Crete)

← IME Keystrokes: .Vv

All of the Unicode values given above should be read as *u+10*nnn; e.g. on the ⌼ key, the values are actually u+1004b and u+10000

Alt+[space] = 𐄁 u+10100 is a word separator line; Ctrl+[space] = · u+10101 is a word separator dot; [no key] × u+10102 is an Aegean check mark.

Aegean Numbers go from 1 to 90999 (no zero); symbols are duplicated using the same pattern as the numbers from 1-9.

Aegean Numbers											
10	20	30	100	200	300	1000	2000	3000	10000	20000	30000
10110	10111	10112	10119	1011a	1011b	10122	10123	10124	1012b	1012c	1012d
𐄐	𐄑	𐄒	𐄙	𐄚	𐄛	𐄢	𐄣	𐄤	𐄪	𐄫	𐄬

Aegean Measures								
Weight:				Dry:	Liquid:			
10137	10138	10139	1013a	1013b	1013c	1013d	1013e	1013f
𐆇	𐆈	𐆉	𐆊	𐆋	𐆌	𐆍	𐆎	𐆏

Русский (кириллический) макет клавиатуры

Russian (Cyrillic) Layout

Pronunciation: Russkiy kirillicheskiy maket klaviatury ← ***IME Keystrokes***: Heccrbq (rbhbkkbxtcrbq) vfrtn rkfdbfnehs

Several letters were dropped from the standard Russian alphabet after the 1917 revolution, so if a Cyrillic letter not shown on this layout is encountered in contemporary text, it is possible that the language is not Russian, but possibly – in alphabetic order – Abkhasian, Azerbaijani, Baskhir, Bulgarian, Byelorussian, Chukchi, Chuvash, Kildin Sámi, Macedonian, Moldovian, Nivkhi, Serbian, Tajik, Ukranian, Uzbeki, Yakut, or other.

Similar to how English typewriters used the small letter L as the number one in order to make best use of the limited number of keys available, Russian typewriters used the З (the letter Ze) key for the number three because for the same reason.

The Cyrillic Ё, ё, Й and й characters (on the [~], [`], [Q], and [q] keys) are not composites of Е, е, И and и (on the [T], [t], [B], and [b] keys) with an added diacritic, but are both entered and stored as independent characters.

Soviet-era portable typewriter – suspiciously similar to the West-German Olympia

Русский (кириллический) макет клавиатуры

Russian (Cyrillic) Code Points

Pronunciation: Russkiy kirillicheskiy maket klaviatury ← ***IME Keystrokes***: Heccrbq (rbhbkkbxtcrbq) vfrtn rkfdbfnehs

รูปแบบแป้นพิมพ์ภาษาไทย

Thai (TIS-820-2538) Layout

Pronunciation: Roop baap paan pimp'pasa Thai

← IME Keystrokes: i^xc[[cxhorb,rN4kKkwmp

Similar Glyphs / Letter Forms in the Thai Alphabet with key presses and pronunciation.

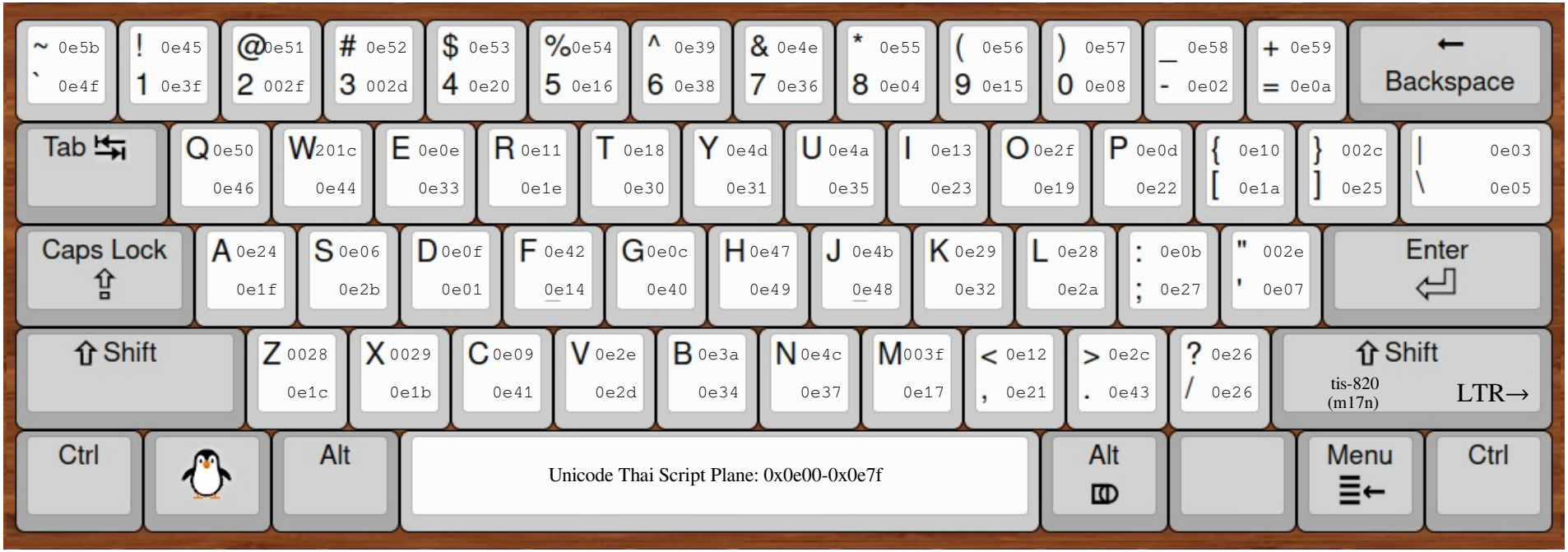
Just as with other layouts provided, Thai numerals are shown in green on the keyboard map above, but do not occupy the same keyboard positions as most other languages and, in fact, are not even contiguous. N.B. Do not confuse the Latin "@" character (shifted 2) with the Thai digit "๑" ("1") (shifted /)!

รูปแบบแป้นพิมพ์ภาษาไทย

Pronunciation: Roop baap paan pimp'pasa Thai

Thai (TIS-820-2538) Code Points

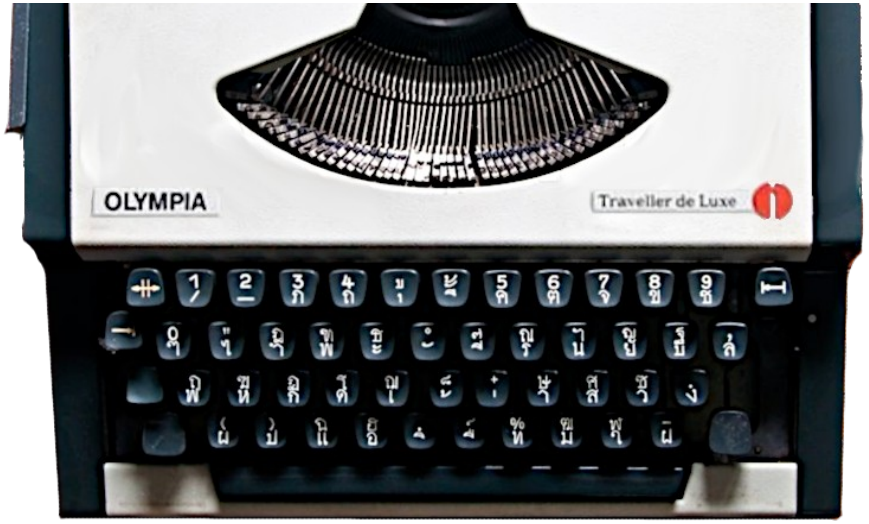
← **IME Keystrokes**: i^xc[[cxhorb,rN4kKkwmp



Except for the use of Arabic number keys on the upper row, the computer keyboard layout specified by the Thai Industrial Standard-820 in the year 2538 (1995 in Western countries) remains essentially the same as it was when the first version of the typewriter shown on the right was introduced in the 1960s.

Most specifically, the physical post-fix "dead keys" (keys that caused no platen advance) are in the same location as the virtual dead keys on modern computer keyboards. Detailed photographs of the heads of several dead keys are provided on page 12 of DDN-2 "Exploring Complex Text Layout."

As with most pre-computer era typewriter keyboards, there are no Latin/English characters on the typewriter.



Türkçe klavye

Pronunciation: tookshay kaláviyay

Turkish (Turkish-F-Latin) Layout

Range: 0100–017F← IME Keystrokes: Hgukbd jlf;d

Prior to 1926, Turkish was written with Arabic Script. Its current Latin alphabet (a b c ç d e f g ğ h ı i j k l m n o ö p r s ş t u ü v y z) doesn't use the letters w, x, and q, but they are still needed for writing words in, or loaned from, English. These letters were therefore 'demoted' to the top row, farther away from the home keys, permitting more efficient typing.

The current official layout has changed very little since computers came into common use, as can be seen in the mid-1970s typewriter layout on the right, and, like other Latin-based typewriters, the small L key served as the digit 1.

A Turkish Pangram – containing all the alphabet's characters (similar to "The quick brown fox…" in English) – is "Pijamalı hasta, yağız şoföre çabucak güvendi." which means "The patient in pajamas quickly trusted the swarthy driver." This can be typed using the keyboard layout above by typing:

Türkçe klavye

Pronunciation: tookshay kaláviyay

Turkish (Turkish-F-Latin) Code Points

← ***IME Keystrokes***: Hgukbd jlf;d

The primary indication that an unknown section of Latin text is Turkish is the presence of capital and small “I” characters both with and without tittles, e.g. I and ı on the R key, as well as İ and i on the S key.

اردو زبان کی بورڈ ترتیب

Urdu (Pakistani, CRULP) Layout

Pronunciation: oardoo keyboard layout

← **IME Keystrokes**: ardw zban ki brwD trtib

Similar Glyphs / Letter Forms in the RTL Urdu Arabic Abjad

N.B. Numeric Characters in Right-to-Left Languages are laid out in Left-to-Right order.

ی	ې	ئ	ي	ب	ت	ث	ح	ج	خ	س	ش	ص	ض	ط	ظ	ع	غ	ف	ق	ك	ل	ن	ه	ة	لأ	لآ	لا
0649	0620	0626	064a	0628	062a	062b	062d	062c	062e	0633	0634	0635	0636	0637	0638	0639	063a	0641	0642	0643	0644	0646	0647	0629	fefb	fef5	fef7
		U		b	t	C	h	j	K	s	x	S	J	y	V	e	G	f	q		l	n					

The Urdu Language spoken in Pakistan is Indic (and very similar to Hindi), but written right-to-left using Arabic Script, reflecting the Muslim majority in Pakistan. This phonetic keyboard layout is that defined by the Center for Research in Urdu Language Processing (CRULP) and seems to be the standard layout now used throughout Pakistani government, business, and academic institutions. Note the paired delimiter reversals on the [ ], [ ], [ ] and [ ] keys; Urdu doesn't use "regular/Western" parentheses. Also note that the question mark is reversed. Urdu is spoken by 52 million people in India and 14.7 million in Pakistan; an additional 214 million speak Urdu as a second language. Actual Arabic forms of the numbers (١٢٣٤٥٦٧٨٩٠) are given precedence over the "Western" forms but, as with numbers in other right-to-left scripts, are written left-to-right.

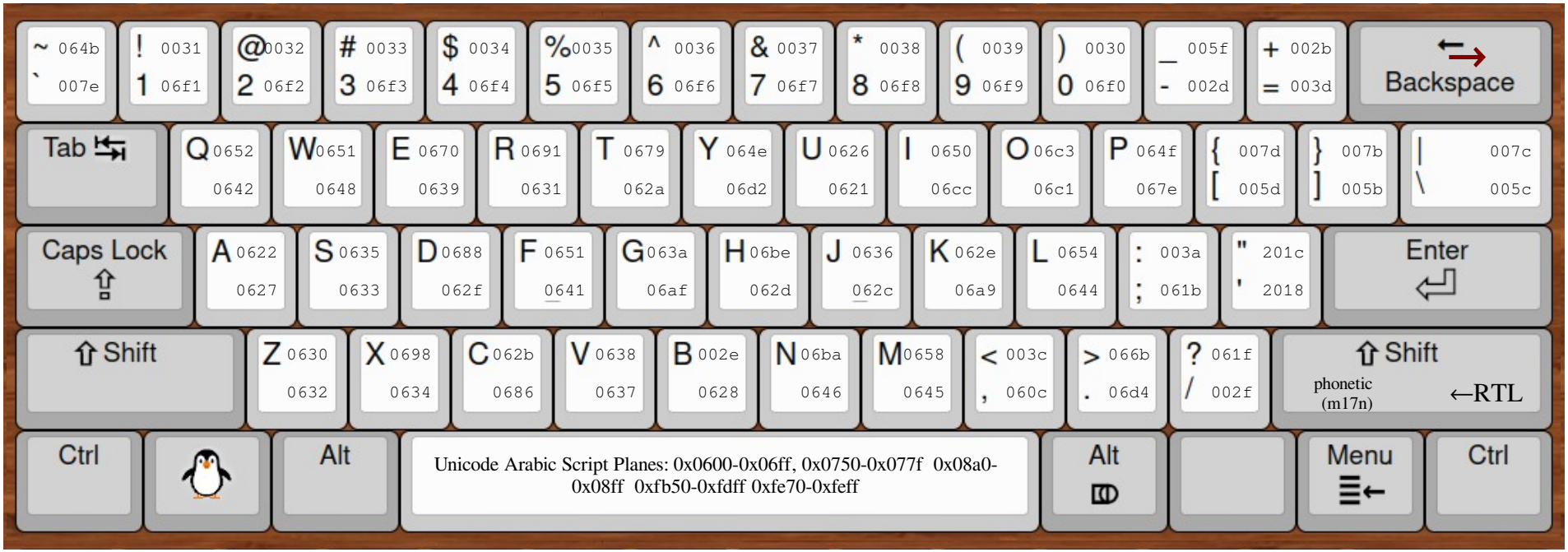
The Anglo-Iranian typographer Sahar Afshar suggests that Urdu character spacing is generally wider than other languages using Arabic script, so the title above uses Khalid Hasny's Amiri font with character spacing set to 0.06pt, since FreeSerif character spacing doesn't expand correctly, though Amiri has deeper line spacing.

اردو زبان کی بورڈ ترتیب

Urdu (Pakistani, CRULP) Code Points

Pronunciation: oardoo keyboard layout

← ***IME Keystrokes***: ardw zban ki brwD trtib



Urdu Text Analysis Example

To determine the key presses required to type زبان – the second word[α] of the title above – copy it into a word processor and move the cursor (▌) back and forth using the arrow keys, noting how, in RTL text, ← is interpreted as **forward** rather than **left**, etc.

e.g. ▌زبان ↔ ن▌زبا ↔ ان▌زب ↔ بان▌ز ↔ زبان▌

There are 4 characters. Now place the cursor after (*left of*) the first character. Use the Word Processor command Alt+x[β], to view the character's value.

The text will change from "بان▌ز" to "U+0632▌بان"; the disconcerting placement of the Unicode value of "ز" from the far right of the word to its left is because the string "U+0632" consists of all Latin LTR characters.

Without repositioning the cursor, press the forward/right arrow to go past the next characters in sequence, pressing Alt+x[β] after each – the end result will be:

U+0632U+0628U+0627U+0646▌ – with all codes in LTR sequence.

Look up each Unicode value in the table's K column: u+632 is a z, u+628 a b, and so forth; therefore زبان can be typed using the z b a n key sequence.

If a particular Unicode value is not in this table (e.g. u+624), the word being analyzed may be from another language that uses Arabic Script for its writing.

Unicode	UTF-8	K
u+60c	D88C	,
u+61b	D89B	;
u+61f	D89F	?
u+621	D8A1	u
u+622	D8A2	A
u+623	D8A3	H
u+625	D8A5	Y
u+626	D8A6	U
u+627	D8A7	a
u+628	D8A8	b
u+629	D8A9	m
u+62a	D8AA	t
u+62b	D8AB	C
u+62c	D8AC	j

Unicode	UTF-8	K
u+62d	D8AD	h
u+62e	D8AE	K
u+62f	D8AF	d
u+630	D8B0	Z
u+631	D8B1	r
u+632	D8B2	z
u+633	D8B3	s
u+634	D8B4	x
u+635	D8B5	S
u+636	D8B6	J
u+637	D8B7	v
u+638	D8B8	V
u+639	D8B9	e
u+63a	D8BA	G

Unicode	UTF-8	K
u+640	D8C0	J
u+641	D8C1	f
u+642	D8C2	q
u+643	D8C3	;
u+644	D8C4	l
u+645	D8C5	m
u+646	D8C6	n
u+647	D8C7	i
u+648	D8C8	w
u+649	D8C9	n
u+64a	D98A	d
u+64b	D98B	~
u+64c	D98C	R
u+64d	D98D	S

Unicode	UTF-8	K
u+64e	D98E	Y
u+64f	D98F	P
u+650	D990	I
u+651	D991	F
u+652	D992	Q
u+654	D993	L
u+658	D997	M
u+66b	D9AB	>
u+670	D9B0	E
u+679	D9B9	T
u+67e	D9BE	p
u+686	DA86	c
u+688	DA88	D
u+691	DA91	R

Unicode	UTF-8	K
u+698	DA98	X
u+6a9	DAA9	k
u+6af	DAAF	g
u+6ba	DABA	N
u+6be	DABD	H
u+6c1	DB81	o
u+6c3	DB83	O
u+6cc	DB8C	i
u+6d2	DB92	y
u+6d4	DB94	.
u+2018	E28098	'
u+201c	E2809C	"

[α] Counting, of course, from the right, as Urdu is written from Right-to-Left.

[β] LibreOffice Writer or Microsoft Word; in Softmaker's TextMaker, use Ctrl+Alt+Shift+X

This section illustrates how to type "Hello, World" in twelve representative languages (plus ancient Greek) on a standard U.S. English ANSI keyboard using the iBus Input Method Editor (IME) with the particular engines listed on the upper right of each section below; other IMEs behave in a substantially similar manner and should produce identical output.

More detailed examinations of five of these Scripts – Arabic, Devanagari, Hebrew, Korean and Thai – are given in Database Design Note 2 (DDN-2) "Exploring Complex Text Layout." Arabic Script is covered more thoroughly in DDN-6 "Exploring Arabic Script Behavior" – The Hànzì Script, used by almost all Chinese dialects and Languages, is covered in DDN-7 "Exploring Han Script." Compare the obvious relationship among the Persian (دنیا), Hindi (दुनिया), Turkish (Dünya), and Urdu (دنیا) words for "world." Also note the common 世界 Hànzì/Kanji syllables that occur in both Chinese and Japanese.

"Hello World" cannot be typed in the ancient Mycenaean Script (Minoan) since only numbers and commercial Linear-B symbols have so far been found and deciphered.

AR — مرحبا بالعالم — l v p f h [ ] f h g u h g l

Arabic – kbd (m17n)
The only explicitly typed space is between the two words (6th character).

ZH — 你好世界 — n i h a o s h i j i e

Chinese – Intelligent Pinyin
Unaccented Latin Pinyin characters are used to produce the written Chinese Hànzì.

FA — سلام دنیا — s g h l [ ] n k D h

Persian – isiri (m17n)
Also known as Farsi or, less often, as Iranian.

EL — Γειά σου Κόσμε — G e i ; a [ ] s o y [ ] K ; o s m e

Greek, Modern (1453-) kbd (m17n)
N.B. ΄ is typed **before** α & ο.

GRC — Χαίρετε Κόσμε — C a i / r e t e [ ] K o / s m e

Greek, Ancient – mizuochi (m17n)
N.B. ΄ is typed **after** ι & ο.

HE — שלום עולם — a k u o [ ] g u k o

Hebrew – kbd (m17n)

HI — नमस्ते दुनिया — v c m d l s [ ] o g v f / e

Hindi – inscript (m17n)
Composite characters formed when typing "md" (स्) and "]s" (ते).

JA — こんにちは世界 — b y i a f Enter p t e [ ] Enter

Japanese Anthy
Kana Typing Mode
Five Hiragana syllables and two Kanji syllables.

JA — こんにちは世界 — k o n n n i c h i h a Enter s e k a i [ ] Enter

Japanese Anthy
Romaji phonetic Typing Mode

KO — 안녕하세요 세계 — d k s s u d g k t p d y [ ] t p r P

Korean – han2 (m17n)
Output adjusts while typing.

RU — Привет, мир — G h b d t n ? [ ] v b h

Russian – kbd (m17n)
Cyrillic Script uses both upper and lower case.

TH — สวัสดีชาวโลก — l ; v l f u = k ; F] d

Thai – tis820 (m17n)
Only the vowel ไ requires use of the Shift key. It is **not** an "upper case" ร consonant.

TR — Selam Dünya — M d l f k [ ] Y g i ; f

Turkish – Turkish
N.B. "u" and "ü" are different; "ü" is not a "u" with an umlaut.

UR — ہیلو، دنیا — O i l w , [ ] d n i a

Urdu – phonetic (m17n)
Linguistically similar to Hindi, but written with Arabic Script rather than Devanagari.

Arabic: arhabana bialealam; Hindi: namaste duniya; Korean: annyeonghaseyo segye; Russian: Privet, mir; Thai: Sawatdee kow lok; Urdu oilo dnia

This test phrase, favored by linguists, is a bit longer than the "Hello, World" usually associated with software development, but provides a more informative example of issues encountered during entry of more complex Scripts.

These examples can be used in the same manner described in the "Hello, World – 14 Examples" section on page 33 for simple system testing. For a better understanding of the processes however, it may be more useful to determine the keystrokes required independently.

Arabic – kbd (m17n) — lugha wāhidah lā takfī

AR

لغة واحدة لا تكف

Chinese Intelligent Pinyin (Mandarin) — yīzhǒng yǔyán yǒngyuǎn bùgòu

ZH

一种语言永远不够

Farsi: aka Persian – Isiri (m17n)

FA

یک زبان هرگز کافی نیست

Greek, Modern (1453-) kbd (m17n) — Mia glóssa den eínai poté arketí.

EL

Μια γλώσσα δεν είναι ποτέ αρκετή.

Greek Ancient – mizuochi (m17n) — Mía glōtta oudepō pote hikanē

GRC

Μία γλώττα οὐδεπώ ποτε ἱκανή

Hebrew – kbd (m17n) — yediat safa ahat eina maspika)

HE

ידיעת שפה אחת אינה מספיקה

Hindi – Inscript (m17n) — ek bhāshā kabhī bhī kāphī nahī hotī

HI

एक भाषा कभी भी काफ़ी नहीं होती

Japanese Anthy – Romaji phonetic genngo wo hitotsu ha ketsu shi te tari nai — Note the mixture of Hiragana and Katagana phonetic "characters"

JA

言語を一つは決して足りない

Japanese Anthy – Kana genngo wo hitotsu ha ketsu shite tari nai

JA

言語を一つは決して足りない

Korean – han2 (m17n) han gaji eon-eoneun chungbunhaji mos hae)

KO

한 가지 언어는 충분하지 못 해

Russian – kbd (m17n) Odnovo yazyka nikogda nedostatochno

RU

Одного языка никогда недостаточно

Thai – tis820 (m17n) paa-săa dieow mâi koie por

TH

ภาษาเดียวไม่เคยพอ

Turkish – Turkish <su ysl fmlf ;dhduls ydeslysu

TR

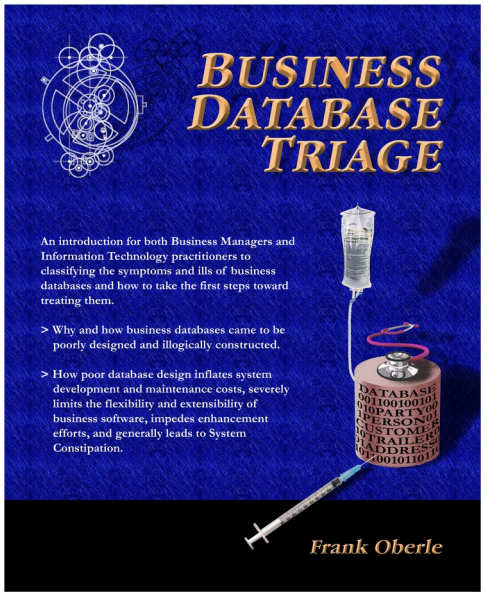
Bir dil asla yeterli değildir

Urdu – phonetic (m17n) ek zabān kabhī bhī kāfi nahīn hotī

UR

ایک زبان کبھی بھی کافی نہیں ہوتی

Antikythera Publications



In addition to an ongoing series of Database Design Notes, Antikythera Publications recently released the book "*Business Database Triage*" (ISBN-10: 0615916937) that demonstrates how commonly encountered business database designs often cause significant, although largely unrecognized, difficulties with the development and maintenance of application software. Examples in the book illustrate how some typical database designs impede the ability of software developers to respond to new business opportunities – a key requirement of most businesses.

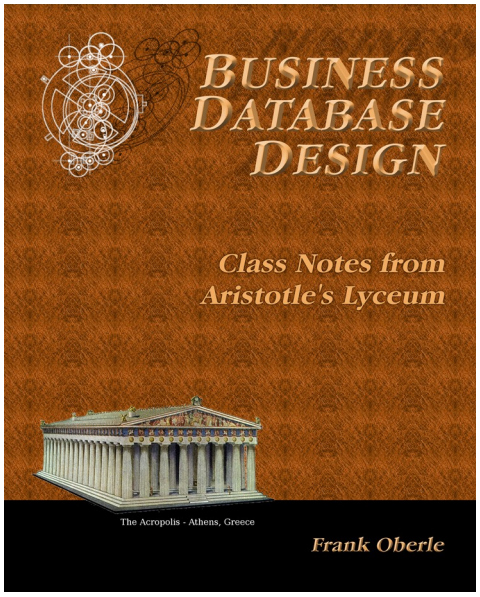
A number of examples of solutions to curing business system constipation are presented. Urban legends, such as the so-called object-relational impedance mismatch, are debunked – shown to be based mostly on illogical database (and sometimes object) designs.

"*Business Database Triage*" is available through major book retailers in most countries, or from the following:

Amazon:

www.amazon.com/Business-Database-Triage-Frank-Oberle/dp/0615916937

More information and sample pages at: www.AntikytheraPubs.com



A follow-up book, "*Business Database Design – Class Notes from Aristotle's Lyceum*" is due to be available in the latter part of 2014.

"*Business Database Design*" leads the reader through the logical design and analysis techniques of data organization in more detail than the earlier work – which concentrated more on understanding and identifying problems caused by illogical database design rather than their solutions.

These logical approaches to data organization, espoused by Aristotle and an "A-List" of his successors, have formed the basis for scientific discovery over more than 2,400 years, and directly led to the technology we deal with today, notably including both relational and object theory.

"*Business Database Triage*" explained the reasons why these principles were virtually impossible to apply during the early years of our transition to the use of computers in business, but since the technology is now sufficiently mature that such compromises can no longer be justified, the time has come to relearn logical data organization techniques and apply them to our businesses.