

This document gives background information on the writing systems for the Right to Left (RTL) languages. Major world languages include Hebrew, Arabic, Farsi, and Urdu.

Bidirectional text

What makes these languages challenging from the perspective of entering text is that while the words are entered Right to Left, numbers are entered Left to Right!! This makes them inherently bidirectional, and results in the need for a cursor that shows what direction you are going at any particular moment. The same applies to mixing languages.



In bidirectional text a distinction may be made between the logical order (how the characters were entered) and the display order (how you see them on the screen). For example, the logical order of a string may be “abc 123 def”. If this is a RTL language it will be displayed “fed 123 cba” starting from the right-hand side of the screen.

Some text editors, including MS Word, will scroll the cursor according to the logical order. This causes a jump when the border between LTR and RTL text is reached, and can be rather confusing and difficult to work with. In an optimal text entry solution, a left arrow press should always move left and right arrow press should always move right. “Clear” Key on a phone or Backspace should always delete opposite the direction the cursor is pointing.

Hebrew terminal letters

There are five letters in Hebrew that take a different shape when they are at the end of a word: kaf, mem, nun, pe, and tsadi.

כ	ך	מ	ם	נ	ן	פ	ף	צ	ץ
Kaf/final kaf		mem/final mem		nun/final nun		pe/final pe		tsadi/final tsadi	

Example:

Hebrew	Transliteration
נן	“nun” (the letter N)

Remember to read the Hebrew starting on the right. Therefore the nun at the right is regular but the nun at the left is in final form. The Unicode font lists the regular and final forms as different characters; the final nun is on a different key (PC keyboard) altogether and the form does not change if I add another letter after it. For a mobile text solution, it is optimal to automatically convert the form for the user depending on whether or not there is another letter following these letters.

Arabic script letters

Almost all Arabic script letters take a different shape depending on their position in a word. This is somewhat similar to cursive handwriting in English. Farsi and Urdu also use the Arabic script and differ in the addition of 4-8 letters to the Arabic alphabet.

Each letter can have up to four shapes: standalone, initial position, medial position, and terminal position. Several letters do not connect to the letters on their left, so they have only standalone and final forms: alef, dal, Dhal, ra, zay, and waw. The example below of building my name shows how letters change shape and the letters alef and ra are limited in their connections.

Example:

Arabic Transliteration

ماریا “Maria”

Built with letters meem, alef, ra, yeh, alef (logical order):

م + ا + ر + ي + ا

م	enter the letter meem
ما	as we enter alef, the meem connects to it
مار	here we enter ra, and we see that alef and ra do not connect
ماری	we enter the letter yeh, and we see that ra does not connect
ماریا	as we enter alef, the yeh connects to it

The Unicode font changes the shape of the letters appropriately as I continue to enter letters, or delete. A mobile text solution must also do this.

Ligatures

There are many ligatures in Arabic, but the critical ligature to cover is lam-alef (“la”) because this is the word “no” in Arabic.

Lam: ل Alef: ا

Standalone/initial position shape: لا medial/terminal position shape: لا

Numbers

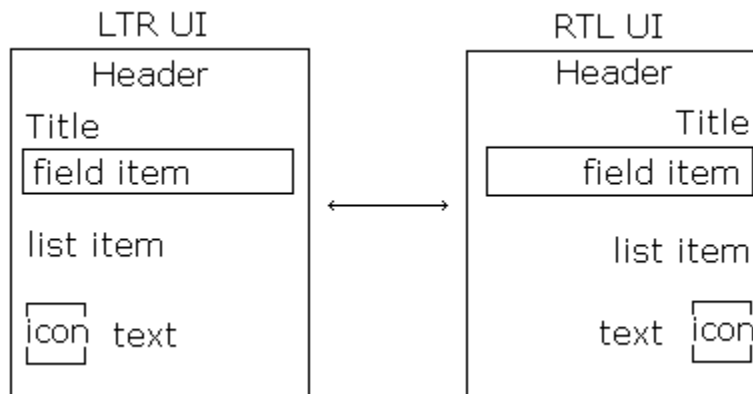
While we may call our numbers “Arabic numerals”, Western numbers are in fact descendants of a combination of Hindi-Arabic digits. Arabic uses different numbers than we do in English.

١٢٣٤٥٦٧٨٩٠
1 2 3 4 5 6 7 8 9 0

Farsi and Urdu digits vary slightly from Arabic, but the Arabic numbers are recognizable. Hebrew uses the same numbers as English.

Alignment

A general rule for switching the UI layout from LTR to RTL is that the left-aligned text is merely changed to right-aligned. Any time an icon is paired with text, it should also be realigned. If a grid highlights the top left cell by default, this should change to top right cell. There need not be changes to items such as status area or soft keys on a keypad phone, and anything centered should remain as is.



Note that in describing the Arabic letters using English we are approximating the sound they make. For this reason transliterations from Arabic to English can sometimes vary, with different English spellings to represent the same sound (i.e. alef vs. alif).

Text examples were built using the Unicode font that can be installed from MS Windows. Other fonts may behave differently. To install fonts: go to Control Panel>Regional Language Options>Languages tab. Choose “Install files for complex script and RTL languages”.

For more information on the RTL languages, I highly recommend www.omniglot.com. This is where I typically begin my research for any language.

- <http://www.omniglot.com/writing/arabic.htm>
- <http://www.omniglot.com/writing/hebrew.htm>
- <http://www.omniglot.com/writing/persian.htm>
- <http://www.omniglot.com/writing/urdu.htm>