

June 2015

Bi-directional (BiDi) implementation guidelines

Developer guidelines

*All implied warranties, including without limitation the implied warranties of merchantability or fitness for a particular purpose, are excluded. In no event shall Sony Mobile or its licensors be liable for incidental or consequential damages of any nature, including but not limited to lost profits or commercial loss, arising out of the use of the information in this document.

SONY

© Copyright 2015 Sony Mobile Communications AB. All rights reserved.
Xperia is a trademark of Sony Mobile Communications AB. Other brands, company or product names mentioned herein are trademarks of their respective owners. You are hereby granted a limited license to download and/or print a copy of this document for personal use. Any rights not expressly granted herein are reserved.

First edition (June 2015)

This document is published by Sony Mobile Communications AB, without any warranty*. Improvements and changes to this text necessitated by typographical errors, inaccuracies of current information or improvements to programs and/or equipment, may be made by Sony Mobile Communications AB at any time and without notice. Such changes will, however, be incorporated into new editions of this document. Printed versions are to be regarded as temporary reference copies only.

Preface

Sony Xperia™ Developer World

For the latest technical news, tutorials and development tools, go to developer.sonymobile.com.

About this document

This document describes guidelines for implementing bi-directional language support. To use this document to best effect you should be familiar with Java programming in an Android environment.

Document history

Change history

2015-06-01	Version 1.0.0	First version
------------	---------------	---------------

Contents

About this document	3
BiDi overview	6
BiDi Languages.....	6
<i>Differences between Arabic, Farsi and Hebrew</i>	<i>6</i>
<i>Right-handed use versus left handed-use</i>	<i>6</i>
<i>Numbers</i>	<i>7</i>
Checklist.....	7
Lists	7
Scroll bars	8
Check boxes	9
Progress bars	10
Pop-up boxes.....	11
Grids.....	12
Flipping	13
Numbers.....	13
BiDi and digits	14
Arabic-Indic digits	14
Eastern Arabic-Indic	14
BiDi languages	15
Arabic.....	15
<i>The Arabic dialects and the modern standard Arabic (MSA)</i>	<i>15</i>
Farsi.....	15
Hebrew	15
Chat languages	16
<i>Chat Arabic</i>	<i>16</i>
<i>Farsi chat</i>	<i>18</i>
<i>Hebrew chat</i>	<i>19</i>
BiDi locales.....	20
General background	20
Calendars.....	20
<i>Calendars in Arabic localisation</i>	<i>20</i>
<i>The Persian calendar</i>	<i>22</i>
<i>The Jewish calendar.....</i>	<i>23</i>
Sorting/Indexing (mixed scripts).....	24
Arabic.....	24
Farsi	27
Hebrew.....	30
BiDi punctuation	33
Brackets.....	33
BiDi reading and writing order	34
Global reading order versus local reading order	34
<i>Numbers in BiDi reading and writing order</i>	<i>34</i>
BiDi scripts.....	36

Arabic.....	36
<i>The basic Arabic alphabet.....</i>	36
<i>Additional Arabic letters.....</i>	39
<i>Ligatures.....</i>	40
<i>Vowels, vowel characters and other diacritics.....</i>	41
<i>Calligraphy, fonts and different writing styles.....</i>	42
Farsi.....	43
<i>The Farsi additions.....</i>	43
<i>Variations from Arabic.....</i>	43
Hebrew.....	44
<i>The basic Hebrew alphabet.....</i>	44
<i>Hebrew vowel points.....</i>	45
BiDi text rendering.....	46
BiDi text input.....	48
General background.....	48
Keyboard layouts and desired interaction.....	48
Symbol input (direction).....	48
Cursor behaviour.....	48
BiDi UI mirroring.....	50
Introduction.....	50
Disclaimer – dependent on context.....	51
To mirror or not to mirror?.....	51
<i>Musts.....</i>	52
<i>Must NOTs.....</i>	52
<i>Desirables (recommended).....</i>	52
<i>Optional (recommended).....</i>	52
UI elements/components.....	53
<i>Media players.....</i>	53
<i>Lists.....</i>	54
<i>Checkboxes.....</i>	55
<i>Progression.....</i>	56
<i>Keyboard arrows.....</i>	57
<i>Icons.....</i>	57
<i>Scrollbars.....</i>	58
<i>Text controls.....</i>	59
<i>“Flipping through” a main menu, home screen.....</i>	60
<i>Tabs.....</i>	61
<i>Date order.....</i>	61
<i>Lock screen.....</i>	61
<i>Status bar.....</i>	61
<i>Date & Time picker.....</i>	61
<i>Theme picker.....</i>	61
<i>Wallpaper picker.....</i>	62
<i>Sound picker.....</i>	62
<i>Music picker.....</i>	62
More information.....	62

Introduction

This document contains guidelines for the implementation of bi-directional languages, such as Arabic, Farsi and Hebrew, in mobile devices.

The intended audience is designers and developers implementing mobile solutions for users of such languages.

These guidelines are originally written for Android based devices but could apply to any mobile phone operating system.

BiDi overview

BiDi is short for bi-directional, it relates to the use of text and numbers in languages that use text written from right-to-left, like Arabic. The bi-directionality describes the way that some of these languages mix text written from right-to-left with numbers and text written from left-to-right. A common reason for this is the integration of English words or slang into other languages.

BiDi Languages

Some examples of BiDi languages are Arabic, Farsi and Hebrew. Those are the BiDi languages currently available in Sony phones and the ones we will concentrate on in this document. It may be useful to note that there are other BiDi languages, such as Urdu, Pashto and Kurdish (when written with the Arabic script).

Differences between Arabic, Farsi and Hebrew

It can be worth noting that the Arabic script is used for Farsi, as well as for Arabic. The difference being that Farsi uses extended letters. This can be viewed in the same way that English uses Latin script, whereas Swedish uses an extended Latin script including additional letters, such as Å, Ä, Ö. This difference in language could affect text input layouts whereas the UI shouldn't be affected.

Regarding UI, Arabic, Farsi and Hebrew appear in similar ways. Small differences may occur, however if nothing specific is mentioned, you can assume that all three languages share the same solution. Note that this does **not** apply to text input.

Right-handed use versus left handed-use

The UI of a BiDi-language is generally mirrored and right-aligned. This means that lists, icons, and other layout components end up on the right side of the screen, which is what local users would expect. Whether a person is right-or-left-handed is never taken into account.

Numbers

The Arab world, for example, can use a mixture of western digits (actually called Arabic numerals) as well as the localised Arabic-Indic digits. Because of this, a setting is absolutely necessary for Arabic speaking users to choose which numeral system they want to use. Importantly, once a user has decided on a numeral system, it will affect all areas where numbers are used, such as the clock, dates and phone numbers. This assumes that such a setting exists and the localised numbers are working in each of those situations.

Numbers in the Arab world are written from left-to-right, the same as western digits. It should be noted that the Arab script in most cases flows from right-to-left. That is the "global" BiDi reading and writing order.

Below are some examples of how numbers are used.

In English	In Arabic (localised digits)	In Arabic (western digits)
1982	١٩٨٢	1982
1983-1987	١٩٨٣-١٩٨٧	1987-1983
1-5	١-٥	5-1
+33 1313131	+ ٣٣ ١٣١٣١٣١	+33 1313131

Note that the Arabic versions are always right-aligned.

Checklist

If you are a designer or developer, consider implementing bi-directional support for the following UI components:

- Lists
- Scroll bars
- Check boxes
- Progress bars
- Pop-up boxes
- Grids
- Flipping
- Numbers
- Players

Lists

Lists are mirrored and right-aligned for right-to-left languages.

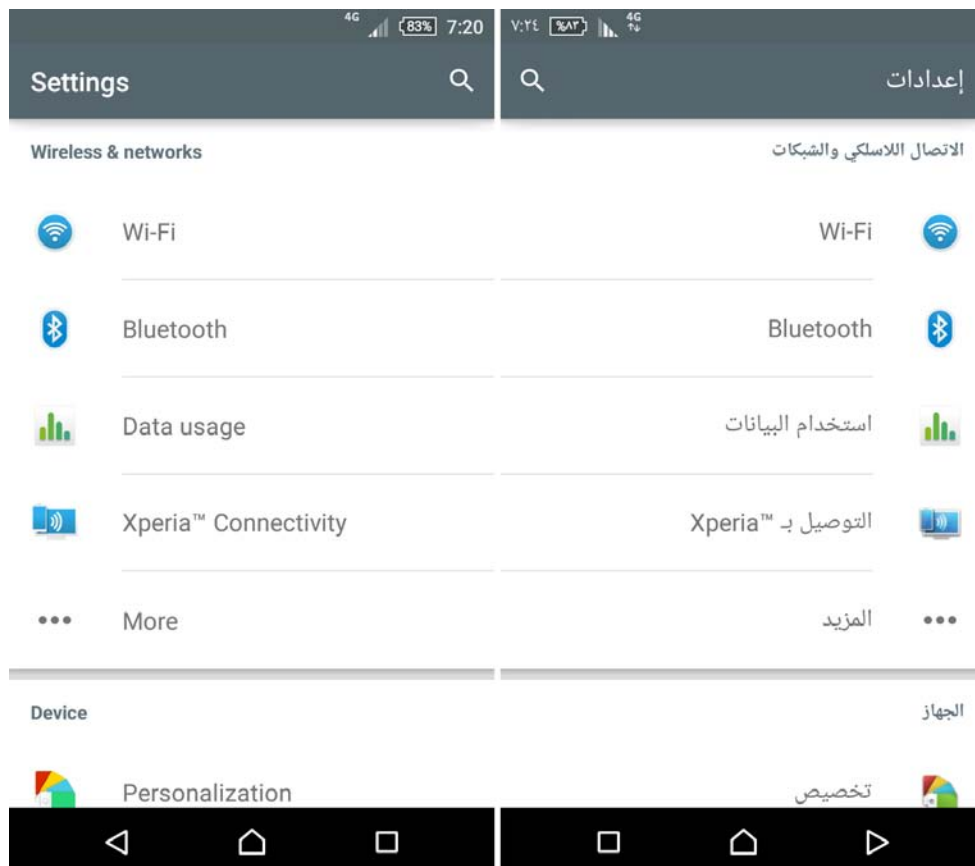


Figure 1. Example of lists in English and Arabic with bi-directional support.

Scroll bars

Scroll bars are mirrored, and should be on the left side for left-to-right languages.

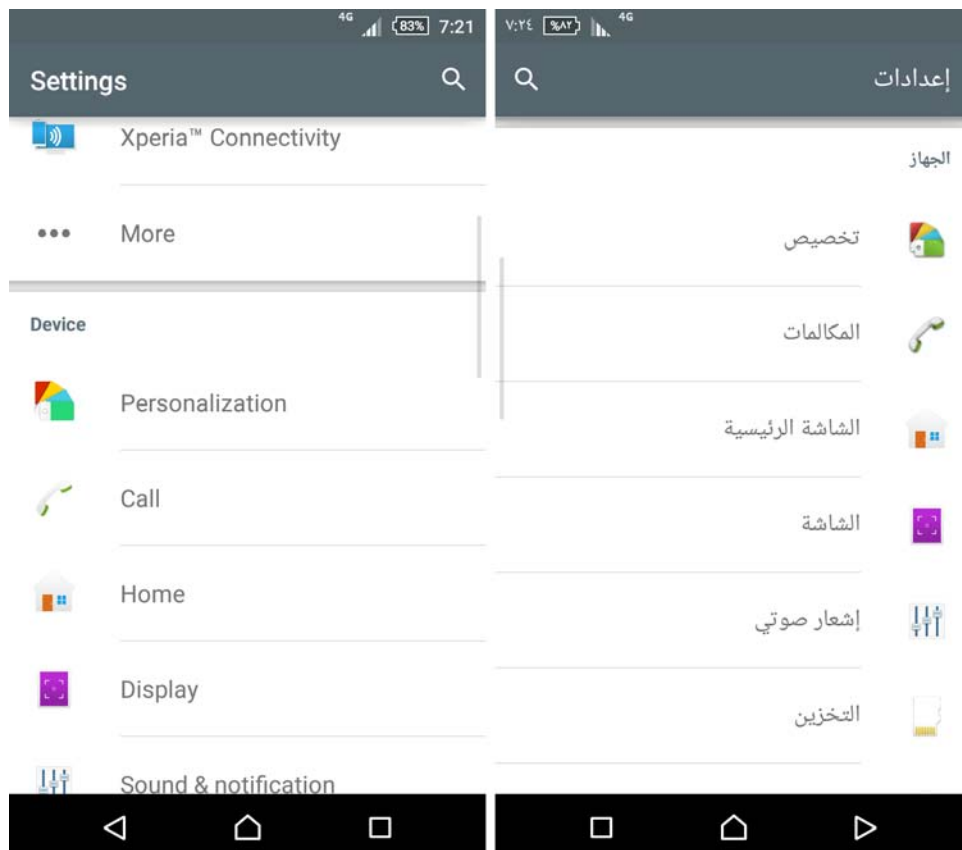


Figure 2. Example of scrolling in English and Arabic with bi-directional support.

Check boxes

Check boxes are mirrored and right-aligned for left-to-right languages. Note that it's not necessary to mirror the actual “check” symbol.

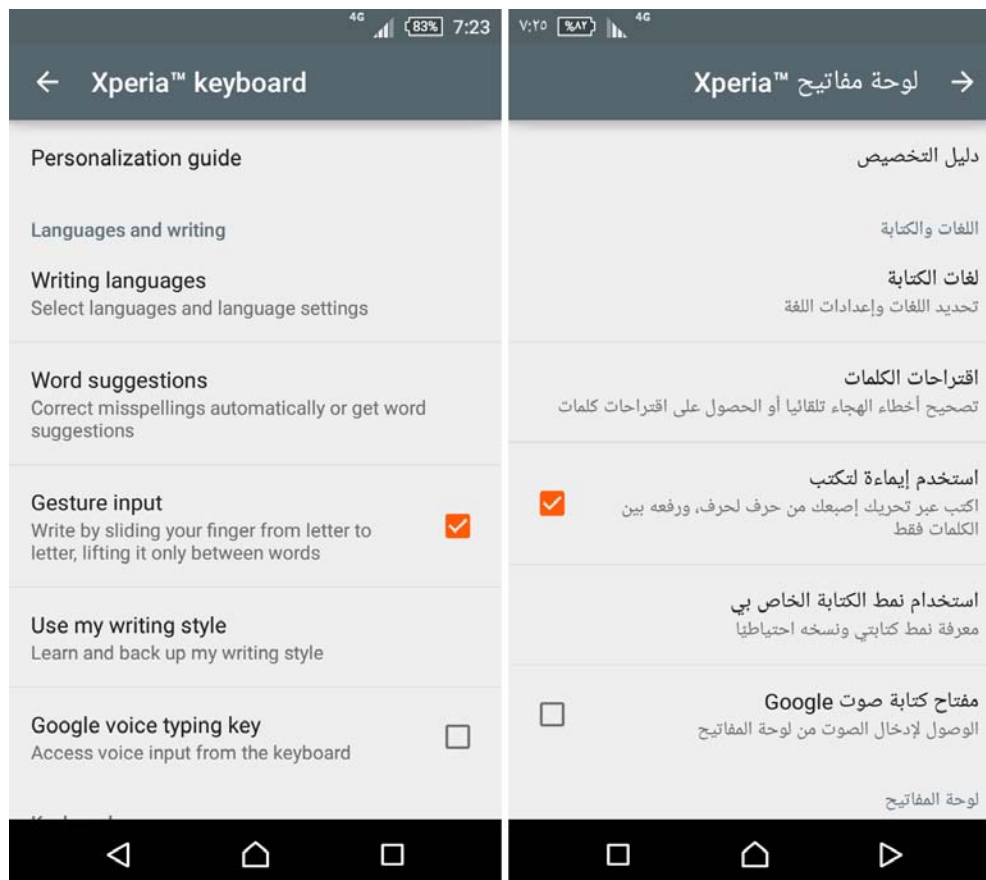


Figure 3. Example of check boxes in English and Arabic with bi-directional support.

Progress bars

Progress bars should be mirrored. This includes progress bars like sound, or when copying files, or backing up or copying contacts from SIM to device. Progress bars that belong to a media player (Music/Video) are not mirrored, to avoid confusion with the PLAY/FORWARD/REWIND buttons.

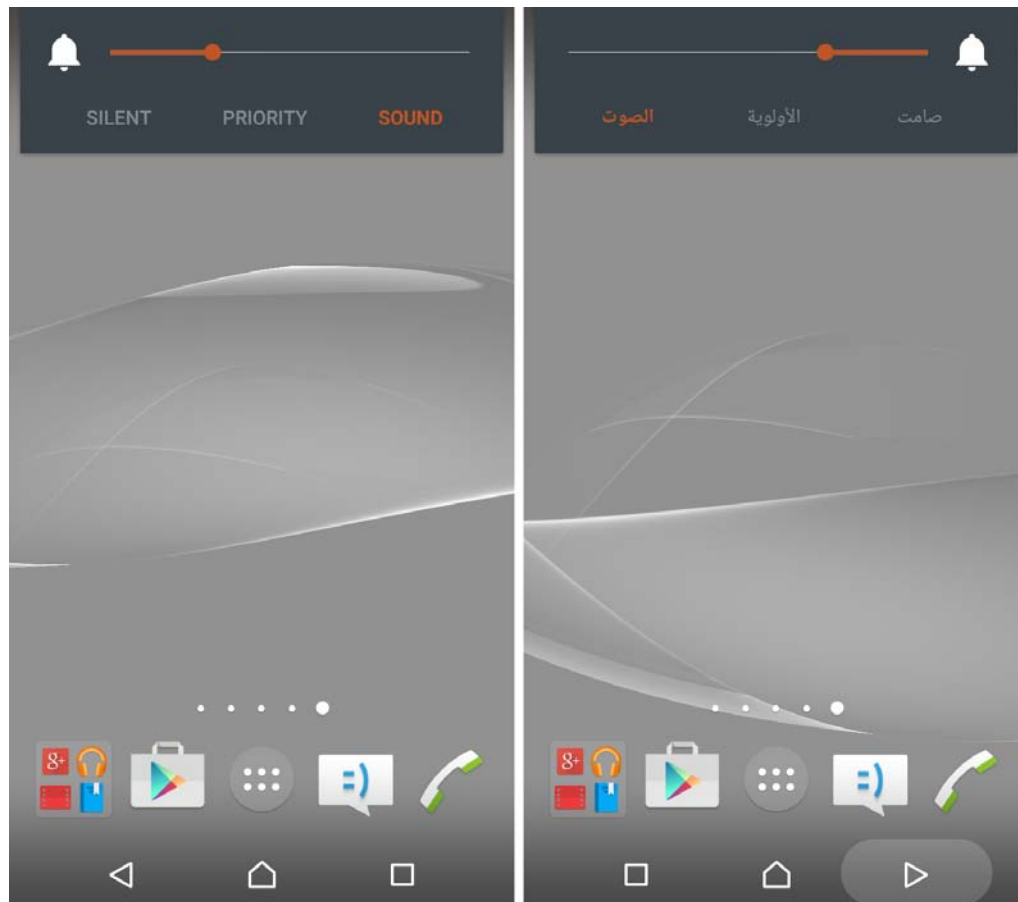


Figure 4. Example of progress bars in English and Arabic with bi-directional support.

All other relevant UI items, text and icons, should be right-aligned. If you are using the Arabic numbers, the [current position] / [total duration] should look like this: [total duration] \ [current position] (also the same if you are showing: "1 out of 4": 1/4 will be 4\1. This is valid for Arabic and Farsi, while Hebrew uses western numbers.

Pop-up boxes

Pop-up boxes buttons are mirrored.

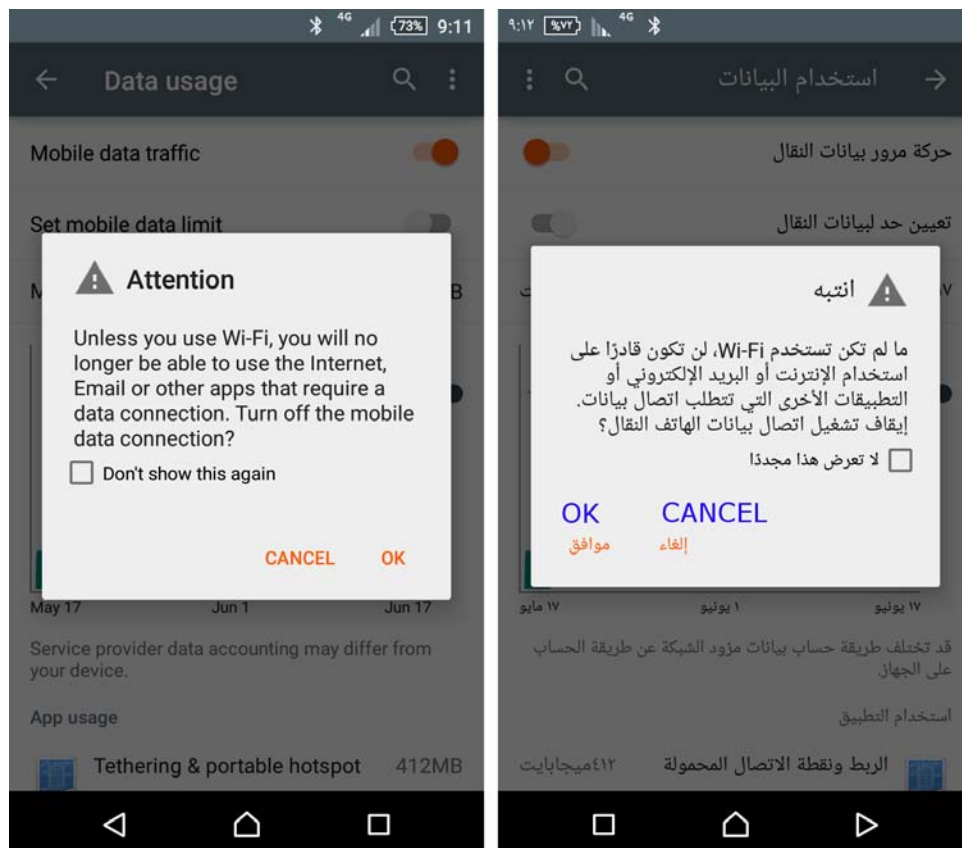


Figure 5. Example of pop-up boxes in English and Arabic with bi-directional support.

Grids

Grids should be mirrored, and should move from right to left for left-to-right languages.

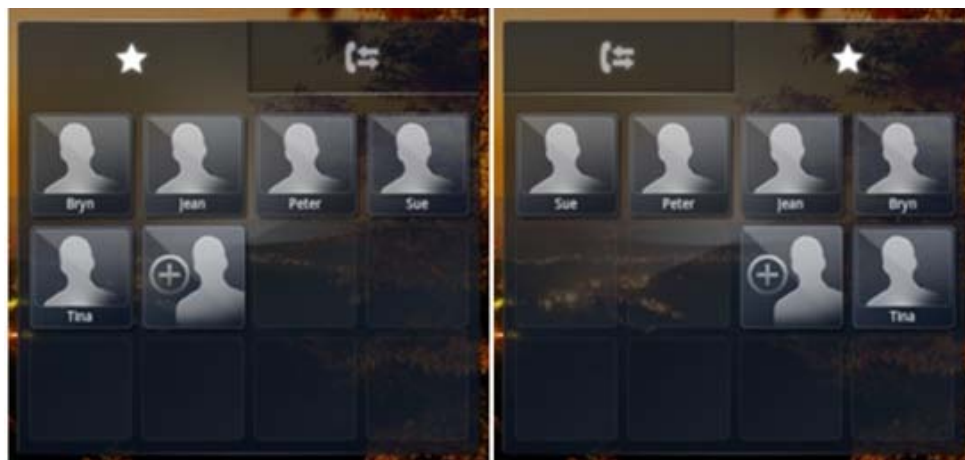


Figure 6. Example of grids in English and Arabic with bi-directional support.

Flipping

The **flipping through** of pictures and albums should be mirrored for left-to-right languages.



Figure 7. Example of flipping in English and Arabic with bi-directional support.

Numbers

Numbers are written from left-to-right. They can often be mixed with text written from right-to-left.

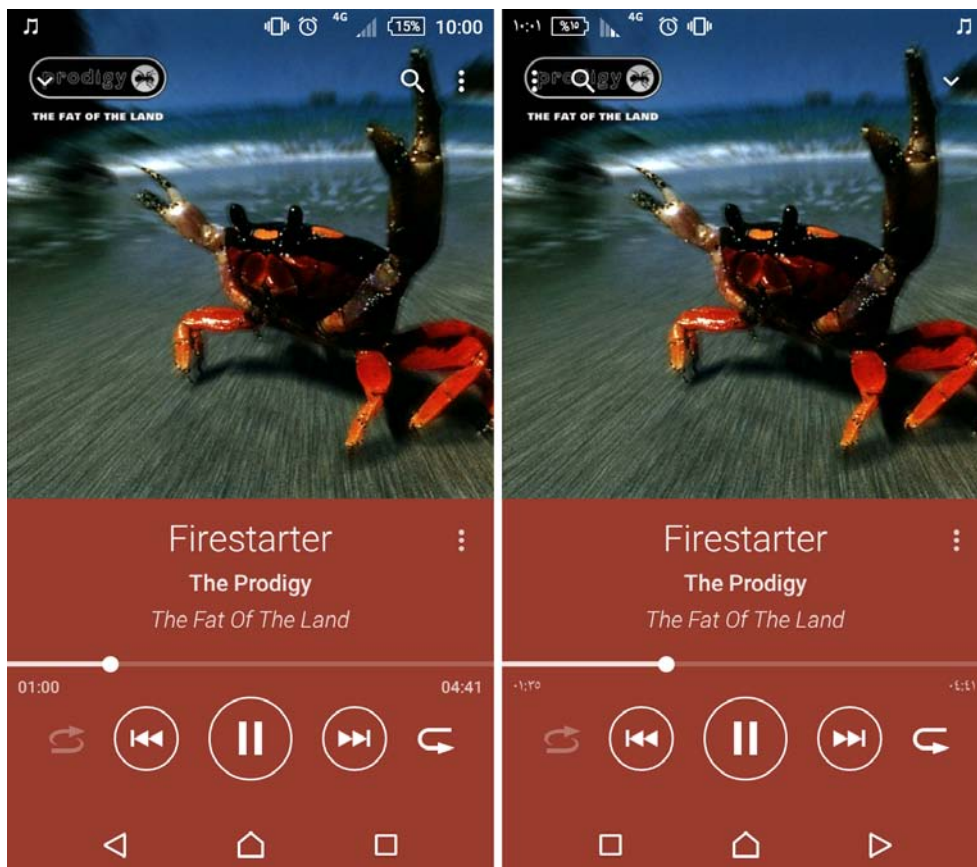


Figure 9. Example of media players in English and Arabic with bi-directional support.

Note that Music and Video players are the only place where the progress bar is **not mirrored**. This ensures consistency with the PLAY/FORWARD/REWIND buttons.

BiDi and digits

Arabic-Indic digits

The Arabic-Indic digits (the localised ones for parts of the Arab world) range between the Unicode values of U+0660 and U+0669.

Arabic-Indic digits	٠	١	٢	٣	٤	٥	٦	٧	٨	٩
Unicode value	0660	0661	0662	0663	0664	0665	0666	0667	0668	0669
"Western" equivalents	0	1	2	3	4	5	6	7	8	9

Eastern Arabic-Indic

In Iran, a modified version of the Arabic-Indic digits is used – they range between the Unicode values of U+06F0 and U+06F9.

Eastern Arabic-Indic digits	۰	۱	۲	۳	۴	۵	۶	۷	۸	۹
Unicode value	0660	0661	0662	0663	0664	0665	0666	0667	0668	0669
"Western" equivalents	0	1	2	3	4	5	6	7	8	9

BiDi languages

Arabic

Arabic is one of the UN's six official languages, it is currently used by about 350 million people in over 22 countries as primary language, and in addition, it is the second official language in a number of other countries, mainly in Africa and Asia.

It is important to note that there is a big difference between the spoken dialects and the written language, which is common to the Arab world.

The Arabic dialects and the modern standard Arabic (MSA)

There are a number of dialects in Arabic which to a large extent is due to the wide geographical spread of the language. Apart from dialects there is also the Modern Standard Arabic (MSA) which is mainly the written form. This is used in education, correspondence, newspapers and literature. The dialects and MSA are often used in different social contexts..

Main geographical areas where these dialects might be used:

- The Arabian peninsula (north and south)
- Syria
- Palestine
- Lebanon
- Jordan
- Iraq
- Egypt
- Maghreb/North Africa (mainly Morocco, Algeria, Tunisia)
- Sudan

Farsi

Farsi is a form of Persian. The term *Persian* covers both Dari (a Persian dialect mainly used in Afghanistan) and Farsi (the Persian used in Iran). Since we want to distinguish between the two, and Sony Mobile at the moment only translates for the Persian used in Iran we use the term 'Farsi'. It is an Iranian language within the Indo-Iranian branch of the Indo-European languages. This means that Farsi is not a Semitic language such as Arabic or Hebrew, which affects grammar, pronunciation and basically all other forms of the language. Farsi (Persian) and its grammar are similar to that of many contemporary European languages.

Hebrew

Modern Hebrew (Ivrit) is the official language of Israel, alongside Arabic – it is a modernized form of Classical Hebrew. Modern Israeli Hebrew is currently spoken by about seven million people - most of them citizens of Israel, or Israelis immigrants living around the world.

Hebrew is written with the Hebrew script.

Chat languages

Chat Arabic

Chat Arabic is sometimes called 'Arabish' (or 'Aralish') and is used to communicate over the Internet or sending text messages using Latin letters and numbers. There are no strict rules for the conversion, and the way of writing varies a lot between dialects and countries. Because of the informal nature of this system, there is no single "correct" way, so some character usage overlaps, for example, 6 is used sometimes for both ط and ح. Regional variants in the pronunciation of the Arabic letter will also affect the character used, for example, ج might be transliterated as j in the Gulf dialect or as g in the Egyptian dialect.

Arabic letters are replaced by letters that are phonetically equivalent, or nearly equivalent, in English. Some letters have no equivalent counterpart and therefore they are represented by numbers or numbers preceded by an accent mark. The most common ones are:

Chat Arabic	Arabic
7	ح
5	خ
9	ص
'9	ض
6	ط
'6	ظ
3	ع
'3	غ
Q	ق
2	ء

The Arabic Chat alphabet is never used in formal settings and is very rarely used for long communications.

See below for a comparison table including the formal transliteration - note that only the most commonly used forms of chat Arabic are shown here:

Chat	Formal transliteration	Arabic
2	'	ء
3	'	ع
gh, 3'	gh, gh	غ
7	h, H, h	ح
kh, 5, 7'	kh, kh, x	خ
S	s, S	ص
D	d, D	ض
T	t, T	ط
Z	z, Z, DH	ظ
2, q	Q	ق
A	a, aa	ا
B	B	ب
D	D	د
d, z	d, z, dh, dh	ذ
F	F	ف
H	H	ه
g, j	j, dj, dj	ج
K	K	ك
l, L	L	ل
M	M	م

Chat	Formal transliteration	Arabic
N	N	ن
R	R	ر
S	S	س
sh, ch	sh, sh	ش
T	T	ت
s, t, th	th, th	ث
w, o, u, ou, oo	w, o, u, ou, uu	و
y, i, ee, e	y, i, ee, e	ي
Z	Z	ز

Farsi chat

Penglish, Pinglish, Fingilish (formed from *Persian + English* or *Farsi + English*) are terms used to describe the way Persian words are written using the Latin alphabet as opposed to the extended Arabic script. This type of writing is commonly used for online, emails and SMS.

Initially there were no strict rules for conversion into Fingilish or how words were written with the Latin script, and the way of writing Fingilish varied greatly by accent. More recently a standard called *Desphilic* script has been introduced, which standardises the concept and usage of a Persian (extended Arabic) -Latin alphabet. *Desphilic* script can be used to officially transliterate Persian and related dialects into a character set suitable for ordinary English keyboards, and also accepts extended Latin characters (Ä, Š, Ž, Ü). *Unipers* is another standard, defining an extended Latin alphabet for the Persian language.

Some examples:

Fingilish	Desphilic	Perso-Arabic (extended Arabic)	Meaning in English
Dorood	DorOd	درود	hello
Amadan	Amadan	آمدن	to come
Parvaz	parvAz	پرواز	structure
Ghamoos	QAmOs	قاموس	manifest

The table below shows how the Latin alphabet is used to write the Persian (Farsi):

Persian (ext Arabic script)	Fingilish	Persian (ext Arabic script)	Fingilish
ا، آ	a (aa)	ص	s
ب	b	ض	z
پ	p	ط	t
ت	t	ظ	z
ث	s	ع، ء	` , ae, ea
ج	j	غ	gh
چ	ch	ف	f
ح	h	ق	q, gh
خ	kh (x)	ک	k
د	d	گ	g
ذ	z	ل	l
ر	r	م	m
ز	z	ن	n
ژ	zh, j	و	u, v
س	s	ه	h
ش	sh	ی	y

Hebrew chat

Hebrew chat is done using both Hebrew letters but also writing it with Latin letters. This is also used in a transliteration form (transliterations of each letter are found under the Hebrew letter table).

BiDi locales

General background

A locale is a geographic region defined by a combination of language and cultural norms. ‘Locale’ is not to be confused with ‘language’. One language, such as French or English, may be used in several countries but each country has its own locale – this includes currency, time and date formats, units etc. With Arabic, which is used in more than 22 countries, a minimum of 4-5 Arabic locales are needed to accommodate different user settings within the Arab world. The following is recommended as a minimum:

- The UAE
- Egypt
- Lebanon or Jordan
- Morocco or Tunisia

Note, if you can only choose one country for the whole Arab world then the best is to choose the UAE.

For Farsi, choose Iran as the natural locale in Xperia phones and for Hebrew, choose Israel.

Calendars

Calendars are normally based on astronomical events, and the two most important astronomical objects are the sun and the moon. The way cycles are calculated can vary depending on the calendar.

Calendars in Arabic localisation

Gregorian calendar with Arabic transliteration

The Gregorian calendar is used in large parts of the Arab world and the month names are mostly written with Arabic letters, even though you can also find them written with Latin letters. When written in Arabic they are mainly based on the English pronunciation, but in Algeria, Tunisia and Morocco you can also find them written based on the French pronunciation.

Month	Arabic transliteration (most common based on English pronunciation)	Transliteration of the Arabic written word to the left *	Arabic transliteration (based on French pronunciation)	Transliteration of the Arabic written word to the left *
January	يناير	<i>yanaayir</i>	جانفييه	<i>shaanfiye</i>
February	فبراير	<i>febraayir</i>	فيفرييه	<i>fifriye</i>
March	مارس	<i>maars</i>	مارس	<i>mars</i>
April	أبريل	<i>abriyl</i>	أفريل	<i>afriyl</i>
May	مايو	<i>maaiyuu</i>	مي	<i>mai</i>
June	يونيو	<i>youniyou</i>	جوان	<i>jouan</i>

Month	Arabic transliteration (most common based on English pronunciation)	Transliteration of the Arabic written word to the left *	Arabic transliteration (based on French pronunciation)	Transliteration of the Arabic written word to the left *
July	يوليو	<i>youliyou</i>	جوييه	<i>jouiye</i>
August	أغسطس	<i>aghustus</i>	أوت	<i>aout</i>
September	سبتمبر	<i>sebtember</i>	سبتمبر	<i>septembre</i>
October	أكتوبر	<i>oktoubre</i>	أكتوبر	<i>oktoubre</i>
November	نوفمبر	<i>noufember</i>	نوفمبر	<i>noufembre</i>
December	ديسمبر	<i>diyseember</i>	ديسمبر	<i>diyseembre</i>

* Due to different transliterations of the Arabic alphabet, other spellings of the months are possible.

Middle Eastern calendar

The month names of this calendar are from ancient origin and some of them are found already in the old Mesopotamian calendar. The calendar is named 'Christian Arabic' by Microsoft but other names may also be found – here we call it Middle Eastern to not confuse it with the Muslim/Hijri calendar which at times is also named 'Arabic'.

The calendar is solar but the month names are completely different from the above although entirely equivalent; for details see below:

Month	Middle Eastern calendar names	English transliteration of the ME calendar names *
January	الثاني كنون	<i>Kanun al-thani</i>
February	شباط	<i>Shubat</i>
March	أذار	<i>Adhar</i>
April	نيسان	<i>Nisan</i>
May	أيار	<i>Ayar</i>
June	حزيران	<i>Haziran</i>
July	تموز	<i>Tammuz</i>
August	أب	<i>Ab</i>
September	أيلول	<i>Aylul</i>
October	الأول تشرين	<i>Tishrin al-awal</i>
November	الثاني تشرين	<i>Tishrin al-thani</i>
December	الأول كنون	<i>Kanun al-awal</i>

* Due to different transliterations of the Arabic alphabet, other spellings of the months are possible

Note – According to Microsoft, the Middle Eastern calendar is the official calendar in Iraq, Syria, Lebanon and Jordan. It is also the second official calendar in all the other Arab countries.

Muslim/Hijri calendar

The Muslim calendar usually consists of six 29-day months and six 30-day months, for a total of 354 days. That's a little more than eleven days short of a solar year. Because of this, all months cycle backwards through the seasons. It is only over a 33 year cycle that lunar months take a complete turn and fall during the same season.

Each new month of the Hijri calendar, is marked by a crescent moon. Muslim months begin at sunset on the day that the lunar crescent is actually sighted, an actual human **physical** sighting of the crescent moon at a given locale.

Since 1999 (Muslim year 1420) the rule has been as follows: On the 29th day of an Islamic month, the times when the sun and the moon set are compared. If the sun sets before the moon, the next day will be the first of a new month; but if the moon sets before the sun, the next day will be the last (30th) of the current month.

The times for the setting of the sun and the moon are calculated for the coordinates of Mecca.

Month number	Muslim month	Transliterated *
1	محرم	<i>Muharram</i>
2	صفر	<i>Safar</i>
3	الأول ربيع	<i>Rabi` al-awal</i>
4	الثاني ربيع	<i>Rabi` al-thani</i>
5	الأولى جمادى	<i>Jumada al-awla</i>
6	الثانية جمادى	<i>Jumada al-thaniya</i>
7	رجب	<i>Rajab</i>
8	شعبان	<i>Sha`ban</i>
9	رمضان	<i>Ramadan</i>
10	شوال	<i>Shawwal</i>
11	القعدة ذو	<i>Dhu-l-Qa`da</i>
12	الحجة ذو	<i>Dhu-l-Hijja</i>

* Due to different transliterations of the Arabic alphabet, other spellings of the months are possible

The Persian calendar

A Persian local calendar exists and below you will find a short overview of that – still, it is mainly the Muslim calendar which is used

The Persian calendar is a solar calendar with a starting point that matches that of the Muslim calendar. Apart from that, the two calendars are not related.

The names and lengths of the 12 months that comprise the Persian year are:

1. Farvardin	(31 days)	7. Mehr	(30 days)
2. Ordibehesht	(31 days)	8. Aban	(30 days)
3. Khordad	(31 days)	9. Azar	(30 days)
5. Mordad	(31 days)	11. Bahman	(30 days)
6. Shahrivar	(31 days)	12. Esfand	(29/30 days)

Due to different transliterations of the Persian (-Arabic) alphabet, other spellings of the months are possible.

The Jewish calendar

As it exists today, the Jewish or Hebrew calendar is a lunisolar calendar based on calculation rather than observation. This calendar is the official calendar of Israel and is the liturgical calendar of the Jewish faith. Note though that the official and the religious calendar starts at different months – the Gregorian calendar transliterated into Hebrew also exists parallel to these.

Hebrew month name
Nisan
Iyar
Sivan
Tammuz
Av
Elul
Tishrei
Cheshvan
Kislev
Shevat
Teved
Adar

The Jewish calendar is counted in 2 ways. In the sacred/religious calendar, Nisan is the first month, but in the civil calendar the New Year (Rosh Hashanah) is celebrated on Tishri 1. This is very important to know because in the *Torah* the months are never referred to by their proper names; instead it mentions the month by number which is the reason for the "Month number" column. They are always referred to as "the first month" (*Nisan*), "the seventh month" (*Tishri*), and so on.

The following is a table comparing the Hebrew/Jewish civil calendar and civil month number with the Hebrew/Jewish religious calendar and religious month number, respectively:

Civil calendar	Month number	Religious calendar	Month number
----------------	--------------	--------------------	--------------

Civil calendar	Month number	Religious calendar	Month number
Tishri	1	Nisan	1
Cheshvan	2	Iyar	2
Kislev	3	Sivan	3
Tevet	4	Tammuz	4
Shevat	5	Av	5
Adar	6	Elul	6
Nisan	7	Tishri	7
Iyar	8	Cheshvan	8
Sivan	9	Kislev	9
Tammuz	10	Tevet	10
Av	11	Shevat	11
Elul	12	Adar	12

Due to different systems of transliteration the Jewish/Hebrew months may be spelled differently.

Sorting/Indexing (mixed scripts)

A general rule for all the BiDi-languages, as well as all other scripts, is that there has to be a functioning mechanism that supports mixing entries of multiple languages. A good example might be songs or contact names in the Phone book/Contacts where names are written in both Latin based letters and with the Arabic script. It is of the utmost importance that these are then sorted or indexed within their respective scripts, and in the Arabic UI Arabic is sorted first then the Latin based entries, in Hebrew the Hebrew shall be sorted first and the Latin-based entries thereafter, and so on. For sorting or indexing within the respective language – see below.

Arabic

How to sort and collate in Arabic is somewhat different depending on where you go in the Arab world, and the principles might differ slightly. It's the same basis for this as with the key print and key mapping, that is, the primary letters follow a natural order but the additional characters such as *hamza* and its variants, *alef* variants as well as *ta marbuta* can be placed in different places in an alphabetical list based on geographical and personal preferences. (See the chapter on [Arabic script](#))

The sort order needs to follow the Arabic (natural) alphabetical order starting with *alef* (0627), and should **not** be sorted based on the Unicode values.

آ	ا	أ	إ	ء	ؤ	ئ
0627	0622	0623	0625	0621	0624	0626

ب	-	-	-	-	-	-
0628						
ت	ة	-	-	-	-	-
062A	0629					
ث	-	-	-	-	-	-
062B						
ج	-	-	-	-	-	-
062C						
ح	-	-	-	-	-	-
062D						
خ	-	-	-	-	-	-
062E						
د	-	-	-	-	-	-
062F						
ذ	-	-	-	-	-	-
0630						
ر	-	-	-	-	-	-
0631						
ز	-	-	-	-	-	-
0632						
س	-	-	-	-	-	-
0633						

ش	-	-	-	-	-	-
0634						
ص	-	-	-	-	-	-
0635						
ض	-	-	-	-	-	-
0636						
ط	-	-	-	-	-	-
0637						
ظ	-	-	-	-	-	-
0638						
ع	-	-	-	-	-	-
0639						
غ	-	-	-	-	-	-
063A						
ف	-	-	-	-	-	-
0641						
ق	-	-	-	-	-	-
0642						
ك	-	-	-	-	-	-
0643						
ل	-	-	-	-	-	-
0644						

م	-	-	-	-	-	-
0645						
ن	-	-	-	-	-	-
0646						
ه	-	-	-	-	-	-
0647						
و	-	-	-	-	-	-
0648						
ي	ى	-	-	-	-	-
064A	0649					

For indexing the basic letters are used (represented by the very left column here above). For full details on collating Arabic – see [Arabic sorting and collating](#).

Farsi

Farsi sorting is slightly different than Arabic. This is not only because of the added letters but also because of how some of the original Arabic letters are used and viewed.

Recommended Farsi sorting order:

ا	-	-	-	-
0622				
آ	-	-	-	-
0627				
ء	أ	إ	ؤ	ئ
0621	0623	0625	0624	0626
ب	-	-	-	-

0628				
پ	-	-	-	-
067E				
ت	-	-	-	-
062A				
ث	-	-	-	-
062B				
ج	-	-	-	-
062C				
چ	-	-	-	-
0686				
ح	-	-	-	-
062D				
خ	-	-	-	-
062E				
د	-	-	-	-
062F				
ذ	-	-	-	-
0630				
ر	-	-	-	-
0631				
ز	-	-	-	-

0632				
ژ	-	-	-	-
0698				
س	-	-	-	-
0633				
ش	-	-	-	-
0634				
ص	-	-	-	-
0635				
ض	-	-	-	-
0636				
ط	-	-	-	-
0637				
ظ	-	-	-	-
0638				
ع	-	-	-	-
0639				
غ	-	-	-	-
063A				
ف	-	-	-	-
0641				
ق	-	-	-	-

0642				
ک	ك	-	-	-
06A9	0643			
گ	-	-	-	-
06AF				
ل	-	-	-	-
0644				
م	-	-	-	-
0645				
ن	-	-	-	-
0646				
و	-	-	-	-
0648				
ه	ه	ة	-	-
0647	06D5	0629		
ی	ی	ي	-	-
06CC	0649	064A		

Hebrew

Hebrew sorting order is as below:

ך	-
05D0	

ב	-
05D1	
ג	-
05D2	
ד	-
05D3	
ה	-
05D4	
ו	-
05D5	
ז	-
05D6	
ח	-
05D7	
ט	-
05D8	
י	-
05D9	
כ	ך
05DB	05DA
ל	-
05DC	

מ	ם
05DE	05DD
נ	ן
05E0	05DF
ו	-
05E1	
ע	-
05E2	
פ	ף
05E4	05E3
צ	ץ
05E6	05E5
ק	-
05E7	
ר	-
05E8	
ש	-
05E9	
ת	-
05EA	

BiDi punctuation

Arabic and Farsi mainly use the same punctuation as the Latin based scripts - such as full-stop, exclamation mark, and colon. There are also a few (namely three) specific punctuation characters as well as a percent sign.

In Arabic	Where and how it is used
؟	Arabic question mark –used exactly as “western” question mark but is just mirrored
،	Arabic comma - used as “western” comma but is “upside down”
؛	Arabic semi colon - used as “western” semi colon but is “upside down”
٪	Arabic percent sign – used as “western” percent sign but the zeros on each side of the slash are only dots since zero in Arabic is depicted with a dot

As mentioned the above is applicable to Farsi too, whereas Hebrew uses the same punctuation as used in the west – **Note:** No mirrored question mark etc in Hebrew.

Brackets

All brackets and parentheses of any sort shall look the same as in Latin script.

BiDi reading and writing order

(See also BiDi text input, in particular, for writing order)

Global reading order versus local reading order

In an Arabic UI (or Hebrew/Farsi) the common reading order is from right-to-left. That is because those languages are written from right-to-left and text flows in that direction. Localised copy might also include left-to-right areas of text or numbers because a Latin script based text, such as an English word, might have been used, hence this text shall be read and written with a "local" reading order from left-to-right.

See examples below:

Text	Comment
أريد أن أكتب رسالة SMS في المحمول	Both the word 'SMS' and in the below text 'GUCCI' are written and read from left-to-right (local order) even though the main text is in Arabic and written/read from right-to-left, i.e. the global order.
معك شنطة GUCCI اليوم	

Numbers in BiDi reading and writing order

The following is an example from the Arabic media. Note the number order:

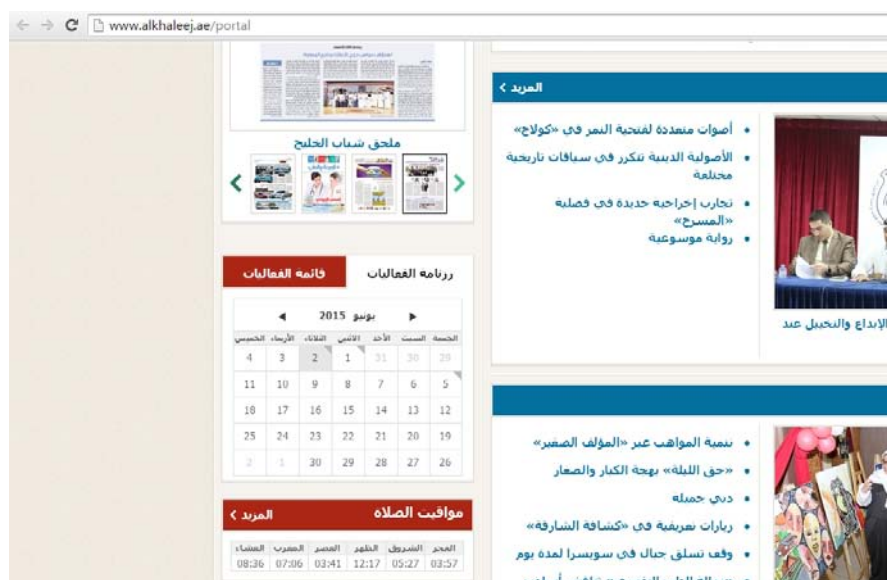


Figure 9. Example of Arabic media.

Date formats

Global and local order –where does one start and the other one stop? Adding numbers which are normally written from left-to-right will add to the complexity – see the following example:

In English	In Arabic (localised digits)	In Arabic ("western" digits)
1982	١٩٨٢	1982
1983-1987	١٩٨٧-١٩٨٣	1987-1983
1-5	٥-١	5-1

In this example digits in one number, such as a year, are written from left-to-right, BUT as soon as any kind of progression comes in the **global** (overall) order takes over, the right-to-left order – hence the 5-1 instead of 1-5 etc.

Adding Arabic (or Farsi/Hebrew) text does not change that. See below:

In English	In Arabic (localised digits)	In Arabic ("western" digits)
September 1982	سبتمبر ١٩٨٢	1982سبتمبر
September 21, 1983-1987	٢١ سبتمبر ، ١٩٨٣ - ١٩٨٧	21 سبتمبر ، 1987 - 1983
September 1 - 5	٥-١ سبتمبر	5-1 سبتمبر

1/3 vs 3\1

As a result of the above the 1/3 (one third) shall in a BiDi languages be written as 3\1 or if with localised digits ٣\١

BiDi scripts

Arabic

Arabic is written from **right to left** and the basic alphabet consists of 29 letters (sometimes 28 is mentioned due to that one of the characters are not considered to be part of the basic alphabet). All letters are consonants but three of them are considered semi-consonants/semi-vowels since they are used to depict long vowels. There are only three vowels; **a**, **u**, and **i**, but there are both short and long ones.

Each basic letter has different forms depending on where in the word the letter is placed, such as if the letter is in the **beginning**, the **middle**, the **end** of a word or placed **independently**. Additionally Arabic is a cursive script, for example, most letters have to connect to each other. On the other hand one does not differ between upper and lower case letters.

The basic Arabic alphabet

The basic Arabic letters with their different contextual forms:

Independent (isolated) letter	Unicode	End form	Middle form	Beginning form	Connection rules	Name and pronunciation
ا	0627	ا	ا	ا	Connects only to the right, with preceding letter.	<i>Alef/ a, aa</i>
ب	0628	ب	ب	ب	Connects both to the right and left, with both preceding and following letter.	<i>Baa/ b</i>
ت	062A	ت	ت	ت	Connects both to the right and left, with both preceding and following letter.	<i>Taa/ t</i>
ث	062B	ث	ث	ث	Connects both to the right and left, with both preceding and following letter.	<i>Thaa/ th, th</i>
ج	062C	ج	ج	ج	Connects both to the right and left, with both preceding and following letter.	<i>Djiym/ dj, dj, j (in Egypt pronounced as 'g')</i>

Independent (isolated) letter	Unicode	End form	Middle form	Beginning form	Connection rules	Name and pronunciation
ح	062D	ح	ح	ح	Connects both to the right and left, with both preceding and following letter.	<i>Haa/ h</i>
خ	062E	خ	خ	خ	Connects both to the right and left, with both preceding and following letter.	<i>Khaa/ kh, kh</i>
د	062F	د	د	د	Connects only to the right, with preceding letter.	<i>Daal/ d</i>
ذ	0630	ذ	ذ	ذ	Connects only to the right, with preceding letter.	<i>Dhaal/ dh, dh</i>
ر	0631	ر	ر	ر	Connects only to the right, with preceding letter.	<i>Raa/ r</i>
ز	0632	ز	ز	ز	Connects only to the right, with preceding letter.	<i>Zaa/ z</i>
س	0633	س	س	س	Connects both to the right and left, with both preceding and following letter.	<i>Siyn/ s</i>
ش	0634	ش	ش	ش	Connects both to the right and left, with both preceding and following letter.	<i>Shiyn/ sh, sh</i>
ص	0635	ص	ص	ص	Connects both to the right and left, with both preceding and following letter.	<i>Saad/ s</i>
ض	0636	ض	ض	ض	Connects both to the right and left, with both preceding and following letter.	<i>Daad/ d</i>

Independent (isolated) letter	Unicode	End form	Middle form	Beginning form	Connection rules	Name and pronunciation
ط	0637	ط	ط	ط	Connects both to the right and left, with both preceding and following letter.	<i>Taa/ t</i>
ظ	0638	ظ	ظ	ظ	Connects both to the right and left, with both preceding and following letter.	<i>Zaa/ z</i>
ع	0639	ع	ع	ع	Connects both to the right and left, with both preceding and following letter.	<i>`Ayn/ `</i>
غ	063A	غ	غ	غ	Connects both to the right and left, with both preceding and following letter.	<i>Ghayin/ gh</i>
ف	0641	ف	ف	ف	Connects both to the right and left, with both preceding and following letter.	<i>Faa/ f</i>
ق	0642	ق	ق	ق	Connects both to the right and left, with both preceding and following letter.	<i>Qaaf/ q</i>
ك	0643	ك	ك	ك	Connects both to the right and left, with both preceding and following letter.	<i>Kaaf/ k</i>
ل	0644	ل	ل	ل	Connects both to the right and left, with both preceding and following letter.	<i>Laam/ l</i>
م	0645	م	م	م	Connects both to the right and left, with both preceding and following letter.	<i>Miym/ m</i>

Independent (isolated) letter	Unicode	End form	Middle form	Beginning form	Connection rules	Name and pronunciation
ن	0646	نـ	نـ	نـ	Connects both to the right and left, with both preceding and following letter.	<i>Nuun/ n</i>
هـ	0647	هـ	هـ	هـ	Connects both to the right and left, with both preceding and following letter.	<i>He/ h</i>
و	0648	وـ	وـ	و	Connects only to the right, with preceding letter.	<i>Waw/ w, ou, uu</i>
ي	064A	يـ	يـ	يـ	Connects both to the right and left, with both preceding and following letter.	<i>Yaa/ y, iy, yy</i>
ء	0621	ء	ء	See below	Special writing rules.	<i>Hamza/ ´</i>

Note that the contextual letters have their own [Unicode values](#). Note that some contextual forms share the same Unicode values. Also note that due to font issues some of the graphical forms herein may appear identical but should differ slightly.

The names of the letters and more importantly how they are transliterated, and when written with Latin letters, this may differ due to the several different systems that are in use. The above system should not be confused with Chat Arabic, which uses numbers and sometimes other Latin letters to depict certain Arabic letters.

Additional Arabic letters

As an addition to the basic letters there are some additional characters which may be characterized as variants of two of the letters above, namely 'alef' and 'hamza'. In addition to those there is a character showing if a word is in the feminine.

Independent (isolated) letter	Unicode	End form	Middle form	Beginning form	Connection rules	Name and pronunciation
أ	0623	أ	أ	أ	Connects only to the right, with preceding letter.	<i>Alef hamza above/ a</i>

Independent (isolated) letter	Unicode	End form	Middle form	Beginning form	Connection rules	Name and pronunciation
ا	0625	ا	ا	ا	Connects only to the right with preceding letter.	<i>Alef hamza</i> below/ i
و	0624	و	و	و	Connects only to the right, with preceding letter.	<i>Waw hamza/</i> short glottal stop where the vowel decides how it shall be pronounced
ي	0626	ي	ي	ي	Connects both to the right and left, with both preceding and following letter.	<i>Ya hamza/</i> short glottal stop where the vowel decides how it shall be pronounced
آ	0622	آ	آ	آ	Connects only to the right, with preceding letter.	<i>Alef madda/</i> a
ى	0649	ى	—	—	Only exits in the end.	<i>Alef maqsura/</i> a, aa
ة	0629	ة	—	—	Only exits in the end.	<i>Taa marbouta/</i> a, -t

Ligatures

Arabic also has a special “ligature” that consists of two characters when written together forming a sign/character of their own. These two letters are lam + alef (ل + ا). A very common word where this can be found is the Arabic word for “No”, ‘laa’ in Arabic لا.

Characters in Arabic (Note! Arabic reading order, right to left)	Unicode (the ligatures have their own contextual Unicode values)	Forms the following ligature in the independent form	Forms the following ligature if something precedes the first letter (‘lam’)
ل + ا	0644+0627	لا	لا
ل + آ	0644+0623	لآ	لآ
ل + ا	0644+0625	لا	لا

Characters in Arabic (Note! Arabic reading order, right to left)	Unicode (the ligatures have their own contextual Unicode values)	Forms the following ligature in the independent form	Forms the following ligature if something precedes the first letter ('lam')
ﻝ + ﺍ	0644+0622	ﻻ	ﻻ

For the specific Unicode values for the contextual ligatures, please see the [Unicode character charts](#)

Vowels, vowel characters and other diacritics

The basic letters shown above are all (in principal) consonants even though three of them function both as (long) vowels and consonants depending on the context. Sometimes they are called semi-vowels or semi-consonants. These three letters are *alef*, *waw* and *ya*. When they function as (long) vowels they are transliterated as such too, for example, long 'a' (aa), long 'u' (uu or ou-you can also see oo) and long 'i' (ii or iy-you can also see yy).

Character (diacritic)	Name and pronunciation of the character	Unicode	How the character is used	Example	Example (translit.)
◌َ	<i>fatha/ a</i>	064E	Is written above the main letter concerned	بَ	ba
◌ِ	<i>kasra/ i</i>	0650	Is written under the main letter concerned	بِ	bi
◌ُ	<i>damma/ u</i>	064F	Is written above the main letter concerned	بُ	bu
◌ْ	<i>sukuun/ no sound</i>	0652	Marks that there is no vowel connected to the main letter concerned	بْ	b
◌ّ	<i>shadda/ double consonant</i>	0651	Marks that the consonant/the main letter concerned is doubled, that is, not only 'b' but 'bb'	بّب	bb
◌ْ◌َ	<i>super script alef</i>	0670	Only used in very few words written above the main letter and pronounced as a long 'a' – in the column to the right here you will see a very common word written with this namely 'Allah'	الله	aa
◌ً	<i>fathatan/an</i>	064B	Is always written in the end – shows that the word is in the accusative.	كِتَابًا	Kitaaban (means 'book')
◌ٍ	<i>kasratan/ in</i>	064D	Is always written in the end – shows that the word is in the genitive.	كِتَابٍ	kitaabin

Character (diacritic)	Name and pronunciation of the character	Unicode	How the character is used	Example	Example (translit.)
ُ	<i>dammatan/un</i>	064C	Is always written in the end – shows that the word is in the nominative basic form	كِتَابُ	<i>kitaabun</i>

Links to Unicode charts - the isolated forms are covered in: [Arabic code charts \(isolated forms\)](#) And most of the contextual forms are covered in: [Arabic code charts \(contextual forms\)](#)

Calligraphy, fonts and different writing styles



Figure 10. Example of Arabic script, al-hurriya (freedom).

In Arabic, the art form of calligraphy is very highly regarded – and you can see one example here above. This has also influenced the modern fonts and it is good to be aware of that letters may look very different from each other depending on which font that is used in a text.

Here are a few examples:

ب	ب	ب	ب	ب
ج	ج	ج	ج	ج
ش	ش	ش	ش	ش
غ	غ	غ	غ	غ
كتاب	كتاب	كتاب	كتاب	كتاب

The font to the very left here is currently very popular whereas the font to the right of that (here in the column marked with a little darker colour) has commonly been used for a long time in longer texts such as books and newspapers.

Farsi

Farsi uses the same script as Arabic (see above) with a few additions.

The Farsi additions

Independent (isolated) letter	Unicode	End form	Middle form	Beginning form	Connection rules	Name and pronunciation
پ	067E	پ	پ	پ	Connects both to the right and left, with both preceding and following letter	<i>Pe/p</i>
چ	0686	چ	چ	چ	Connects both to the right and left, with both preceding and following letter	<i>Tch eh/ ch, tch</i>
ژ	0696	ژ	ژ	ژ	Connects only to the right, with the preceding letter	<i>Jeh/ j, zh</i>
گ	06AF	گ	گ	گ	Connects both to the right and left, with both preceding and following letter	<i>Gaaf/ g</i>

Variations from Arabic

Farsi also modifies some letters from the Arabic alphabet. For example, *alef with hamza below* (ا) changes to *alef* (ا); words using various hamzas get spelled with yet another kind of hamza (so that مسئول becomes مسئول); and *ta marbuta* (ة) changes to *he* (ه) or *ta* (ت). The pronunciation of the Arabic letters is also different.

The letters different in shape are:

Original Arabic character	Farsi modification	Unicode	End form	Middle form	Beginning form	Connection rules	Name and pronunciation
ك	ک	06A9	ک	ک	ک	Connects both to the right and left, with both preceding and following letter	<i>ke/ k</i>

Original Arabic character	Farsi modification	Unicode	End form	Middle form	Beginning form	Connection rules	Name and pronunciation
ي	ی	06CC	ی	ی	ی	Connects both to the right and left, with both preceding and following letter	yeh/y, i, iy

Hebrew

The Hebrew alphabet is used to write the Hebrew language and is written from right-to-left. The alphabet consists of 22 letters; five of them have different forms when they are used at the end of a word which makes five additions to the 22 basic characters and the script also includes diacritics.

The basic Hebrew alphabet

Independent character	Unicode	End form (if applicable)	End form Unicode	Name and pronunciation
א	05D0	—	—	alef/ ´
ב	05D1	—	—	bet/ b, v
ג	05D2	—	—	gimel/ g
ד	05D3	—	—	daled/ d
ה	05D4	—	—	he/ h
ו	05D5	—	—	vav/ v
ז	05D6	—	—	zaiyn/ z
ח	05D7	—	—	het/ h, ch
ט	05D8	—	—	tet/ t
י	05D9	—	—	yod/ y, i
כ	05DB	ך	05DA	kaf/ k, kh
ל	05DC	—	—	lamed/ l
מ	05DE	ם	05DD	mem/ m
נ	05DF	ן	05E0	nun/ n
ס	05E1	—	—	samekh/ s
ע	05E2	—	—	‘aiyn/ `
פ	05E3	ף	05E4	pe/ p, f
צ	05E5	ץ	05E6	tzadi/ z
ק	05E7	—	—	qof/ k

Hebrew vowel points

Vowel point	Unicode	Note	Name and pronunciation
◌◌	05B0	—	<i>Schwa</i> /e, -
◌ֿ	05B1	—	<i>Hataf segol</i> / e
◌ַ	05B2	—	<i>Hataf patah</i> / a
◌ֿֿ	05B3	—	<i>Hataf kamatz</i> / o
◌ִ	05B4	—	<i>Hiriq</i> / i
◌ֵ	05B5	—	<i>Zeire</i> / e
◌ֶ	05B6	—	<i>Segol</i> / e
◌ֹ	05B7	—	<i>Patach</i> / a
◌ֻ	05B8	—	<i>Kamatz</i> / a or o
◌ױ	05B9	—	<i>Holam</i> / o
◌װ	05BB	—	<i>Kubutz</i> / u
◌ױױ	05BC	Same as dagesh below	<i>Shuruk</i> / u
◌ױױױ	05BC	—	<i>Dagesh</i> (<i>stress</i>)

BiDi text rendering

Arabic and Farsi (in particular those two since Hebrew doesn't require connecting) require special rendering since:

1. They are written from right-to-left.
2. The letters have their own contextual forms which have to be connected according to the basic rules of the language.
3. They have diacritics, such as small signs that are written above or under the basic letters (but are only occasionally written out).

Please see the table below for same letters, with and without correct contextual forms. SW algorithm has to handle this correctly otherwise text becomes basically unreadable. This means the contextual Unicode forms shall be supported and used (see above).

Correct	<i>Incorrect</i> (same letters as in the left column here but not connected as they should be)
الهدم	د م ه ل ا
مهندسة	ة س د ن ه م
كتابة	ة ب ا ت ك
جنسية	ة ي س ن ج
ظلمة	ة م ل ظ

The 'harakaat' (short vowels and diacritics) also need special rendering since they are built above or below the basic Arabic (and Farsi) letter. They are small signs written above or below the basic text (but which are only occasionally written out – still, they have to be supported). This means that there need to be space enough in height to add these letters as well as they need to be placed correctly in position of the base letter they “belong” to.

Support the following:

Character/symbol	Name of character/symbol	Unicode
َ	<i>Fatha</i>	064E
ِ	<i>Kasra</i>	0650
ُ	<i>Damma</i>	064F
ً	<i>Fathatan</i>	064B
ٍ	<i>Kasratan</i>	064D

Character/symbol	Name of character/symbol	Unicode
◌̣	<i>Dammatan</i>	064C
◌̣̣	<i>Sukuun</i>	0652
◌̣̣̣	<i>Shadda</i>	0651
-	<i>Tatweel</i>	0640
◌̣̣̣̣	<i>Superscript alef</i>	0670

The special Arabic punctuations marks; Arabic question mark (U+061F), Arabic comma (U+060C), Arabic semicolon (U+061B), and the Arabic percent sign (U+066A) have to be supported.

The *lam-alef* ligatures have to be supported. When the letter *lam* (U+0644) is written together with the letter *alef* (U+0627) they form a special ligature which has its own Unicode, for example, FEFB (when isolated or independent) and FEFC (end form).

All variants of the lam-alef ligature also have to be supported, such as *lam+alef hamza above* (U+0644+0623, and its contextual forms FEF7 and FEF8), *lam+alef hamza below* (U+0644+0625, and its contextual forms FEF9 and FEFA), *lam+alef madda* (U+0644+0622, and its contextual forms FEF5 and FEF6).

BiDi text input

General background

In general, text input offers quite a lot of challenges to a developer, mainly due to the bi-directional nature of the writing. Bi-directional writing is used since Arabic (and other right-to-left languages such as Hebrew and Farsi) are written in a form known as right-to-left (RTL), in which writing begins at the right-hand side of a page and concludes at the left-hand side. This is different from the left-to-right (LTR) direction used by most languages in the world. When LTR text is mixed with RTL in the same paragraph, each type of text should be written in its own direction, which is known as *bi-directional text*. This can get rather complex when multiple levels of quotation are used.

Bi-directionality also applies to mixing RTL languages with LTR numbers.

In Unicode encoding, all non-punctuation characters are stored in writing order. This means that the writing direction of characters is stored within the characters. If this is the case, the character is called "strong". Punctuation characters however (there are only a few localized), can appear in both LTR and RTL languages, and they are then called "weak" characters because they do not contain any directional information. So it is up to the software to decide in which direction these "weak" characters will be placed. Sometimes (in mixed-directions text) this leads to display errors, caused by the BiDi-algorithm that runs through the text and identifies LTR and RTL strong characters and assigns a direction to weak characters, according to the algorithm's rules. To correct or prevent these errors, one may use "pseudo-strong" characters. These Unicode control characters are called "marks". The mark (U+200E LTR or U+200F RTL) is to be inserted into a location to make an enclosed weak character inherit its writing direction. Note though! This may not solve the problem if different technical solutions/platforms are "mixed".

Keyboard layouts and desired interaction

Keyboard layout depends on the a number of factors – language, the “basic” suggested keyboard solution (such as for English/basic Latin), the technical platform (since there might be different limitations to the suggested solutions for Hebrew, Farsi and Arabic) and most importantly, very little standardizations exist, hence it's of the utmost importance to know the background and make sure that Sony Mobile stays consistent with its own previous solutions (if no new standardizations have come up). Localised key board and key mapping specifications for all different products – specific to each product (or platform) - exist.

Symbol input (direction)

The symbols input grid in Arabic, Farsi and Hebrew shall be mirrored and highlight shall be mirrored and **right-aligned**. As all progression in a BiDi language this shall move from right-to-left and begin to the very right of the grid.

Cursor behaviour

In a BiDi UI, the cursor **always has to be positioned to the very right as a default**. Once the first letter has been typed then the cursor should adopt to which input language that is actually used.

Correct cursor behaviour includes support for:

- Bi-directional behaviour (Arabic/Farsi/Hebrew mixed with Latin script and numbers)
- Visual order vs. logical order – when to implement what.

With today's complex text input solutions for English (Latin script), it is sometimes difficult or even impossible at times to find good solutions for the BiDi languages. One has to weigh different user scenarios and come up with solutions that give the least negative impact; one solution might work great in one or two particular interaction situations but will work really badly in another situation – the question then is – is this worth it? Where and how will this impact the user? To be able answer this you need to consult with someone who can weigh the different user scenarios against each other, who has the history, the background and experience. Please check with a UI Localisation expert for the correct feedback and help/information on BiDi text input support.

BiDi UI mirroring

Introduction

As Arabic, Farsi and Hebrew are languages written from right-to-left most of the user interface has to be **mirrored** and **right-aligned**. There are a few exceptions, based on the fact that numbers are written from left to right. These are exceptions, though, and the “beginning” of something is in general to the very right where as for westerners the beginning in general is to the very left.

In general all UI should be mirrored which includes the scroll bar which should be to the left instead of the right, the progress bar should show progression from right to left, lists with icons should be to the very right.

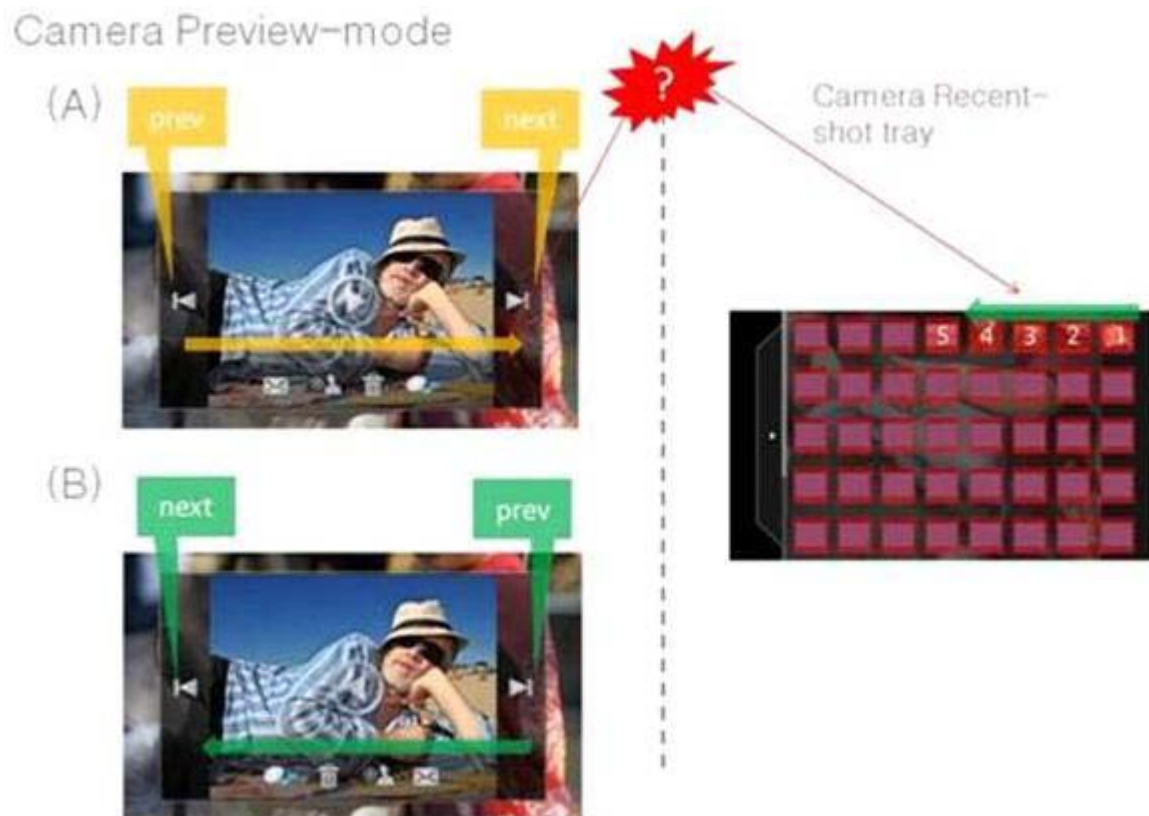


Figure 11. Example of camera UI layout in English (A) and localised BiDi version (B).

In the above example, layout (A) is correct for English and any left-to-right-language and layout (B) exemplifies how it should look in a localised BiDi version. And as the grid to the right shows, this is how it shall look for the BiDi UI.



Figure 12. Example of a localised BiDi UI layout.

Disclaimer – dependent on context

As mentioned above, exceptions to the basic rules may occur and an example of this is, for instance, the play, forward and backward buttons and progress bars in the Players (see below). There may also be certain types of interaction solutions where two principles of the basic localisation rules may collide hence one rule has to be chosen even if it is not perfect but is deemed best under the circumstances. This is about weighing different user scenarios against each other and choosing the most appropriate solution.

To mirror or not to mirror?

There are actually four steps when localising for a BiDi language:

1. The UI you HAVE to mirror
2. The UI you MUST NOT mirror
3. The UI which is desirable to mirror
4. The UI which is optional to mirror.

The two steps are recommended for all developers and designers to prepare for mirroring to keep consistency.

Musts

Any kind of progression HAS to be mirrored since the beginning is considered to be to the very right; calendars (weeks progressing, months, years), progress bars, going from one page to another, scrolling through pictures, or songs. Anything that has a beginning to the left and/or is left-aligned in a left-to-right-language UI should be mirrored and right-aligned in a BiDi language; including lists, lists with icons, tabs, and overlapping menus (so-called cascading menus).

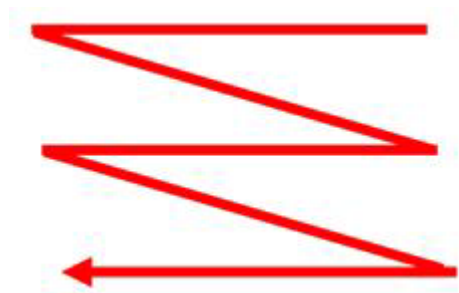


Figure 13. Recommended direction of UI progression.

Must NOTs

PIN numbers or any kind of numeral input shall not be mirrored – right-aligned yes, but not mirrored since digits are written from left-to-right whether “our western” digits or the localised ones are used. Names and other texts not translated shall not be mirrored but shall be right-aligned.

Desirables (recommended)

When there is no clear beginning such as for instance the status bar with the battery icon (see Icons below), the clock etc it's not a must to mirror since there is no absolute beginning. It is desired to do so though and we recommend that all “desirables” are mirrored.

Optional (recommended)

A UI that has two buttons with for instance ‘Save’ and ‘Cancel’ does not have to be mirrored since it does not change anything for the local user – there is no clear beginning. Still, we recommend that all “optional” are mirrored to keep the consistency.

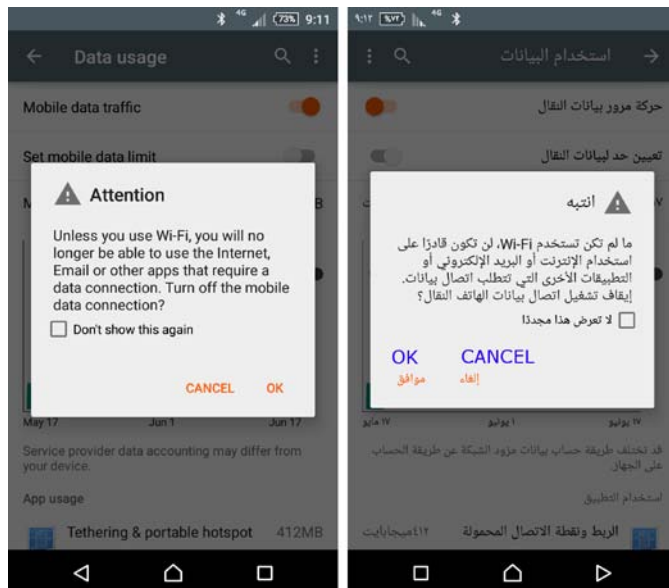


Figure 14. Examples of an English and localised UI with two button options.

UI elements/components

Media players

For media players, (Music and video), the PLAY BACKWARD and FORWARD buttons and the progress bar shall **not be mirrored**, as mirroring the progress bar and keeping the play/backward/forward unmirrored will create lots of inconsistency and strange unexpected problems.



Figure 15. Example of a localised UI for a media player.

Lists

All lists shall be mirrored, and if they are left-aligned in the English UI they shall be right-aligned in the BiDi-language UI. Since lists in most cases are left-aligned in English they shall be right-aligned in a BiDi-language, and by mirroring a UI you will by default create the correct BiDi UI. Example:

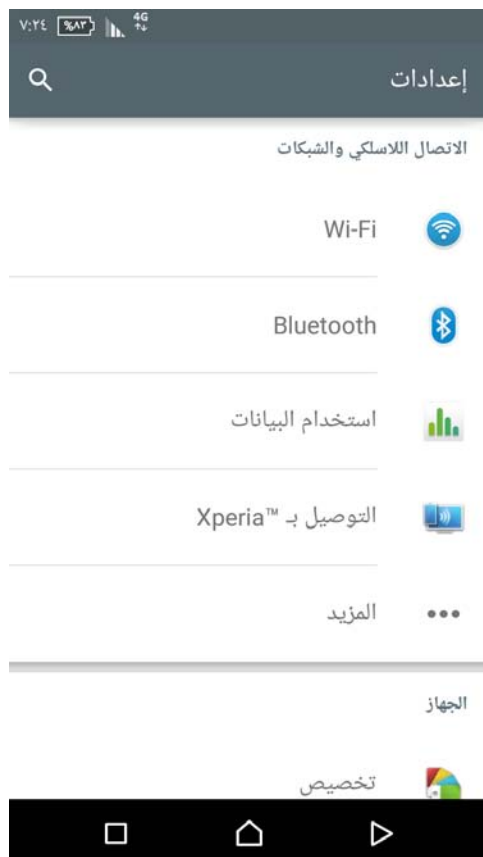


Figure 16. Example of a BiDi UI.

Above a “bidified” UI. (*Note!* English names and texts are not mirrored but written as it should be in English (or any left-to-right-language). That goes for anything not translated. In most cases it will be right-aligned, though.). If a UI is centred in the left-to-right language UI it shall of course remain centred in the bidified UI.

Checkboxes

Checkboxes and UI with check boxes should be mirrored, but the “check” itself shall not be mirrored – we don’t normally do that. It is no disaster if it does happen but should as a rule not be in a BiDi UI.



Figure 17. Example of an English and BiDi UI with checkboxes.

Progression

Progression shall in general be mirrored for a right-to-left language, that is, progression should grow to the left.

Progress bars

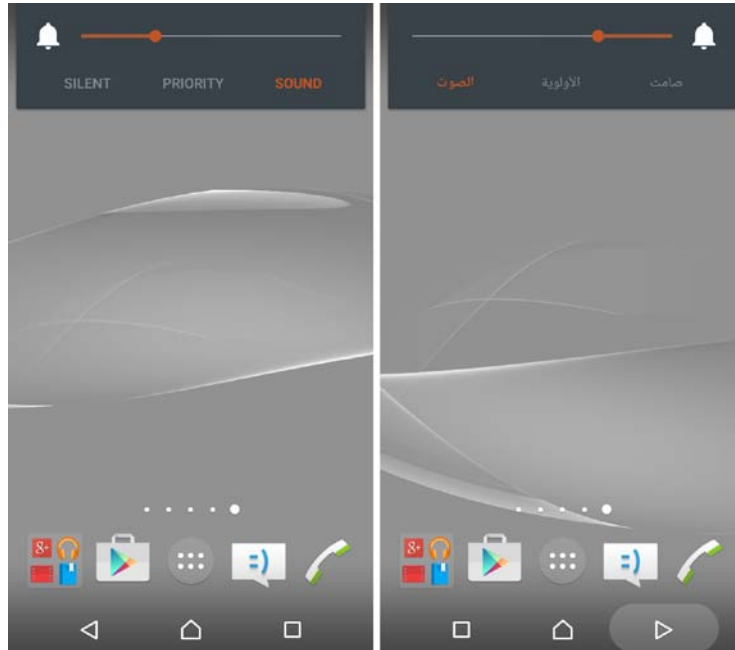


Figure 18. Example of an English and BiDi UI with progress bar.

Keyboard arrows

Keyboard arrows, Backspace, and the Enter key should NOT be mirrored in a BiDi UI. This again goes back to hardware keyboards that were not changed in some of these languages and the local users got used to this user interface and hence today it would be strange to mirror them. Also note that the functionalities of the keys are to remain as they are, such as if I press the key with the arrow pointing right I will expect the highlight/cursor to move right.

Icons

Even though about 90 % of the UI requires mirroring, not a lot of the icons are affected. It is impossible though, to say which icons need to be mirrored and which do not because that depends on the functionality, that is, a co-operation between designers and developers. For example, see picture below:

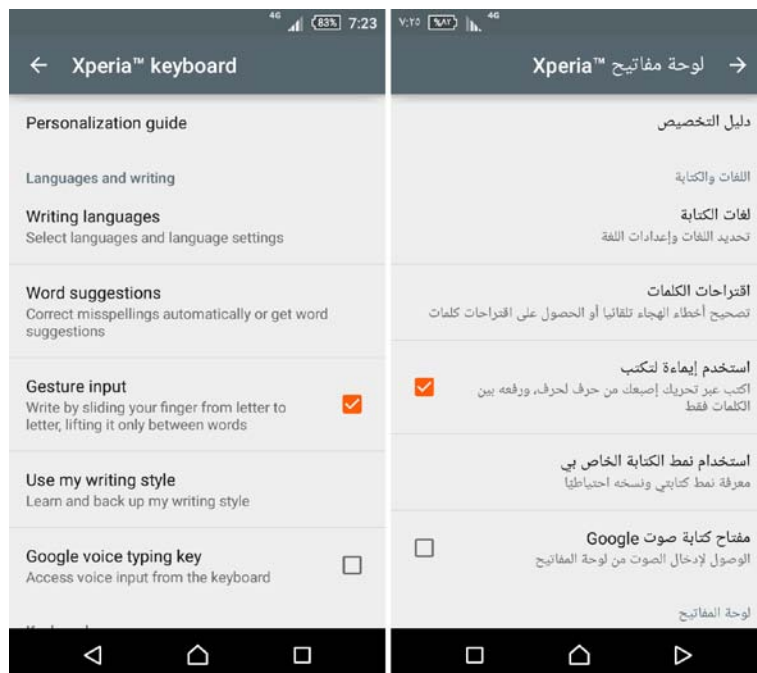


Figure 19. Example of a mirrored UI with Back arrow key (Top left screen in the English UI, going back one step in keyboard setting screen) pointing in the right direction (to the right).

Here, the Back key arrow is pointing to the right which is correct in a mirrored UI since the beginning is to the very right. However it also has to be functionally correct.

Scrollbars

The scroll bar should be positioned to the left, that is, opposite to what it is in a Latin text based UI. If you look at web pages on the Internet you may find certain scroll bars placed in the same way as for English or “western” scroll bars, but again, this can be due to an individual choice of doing so.

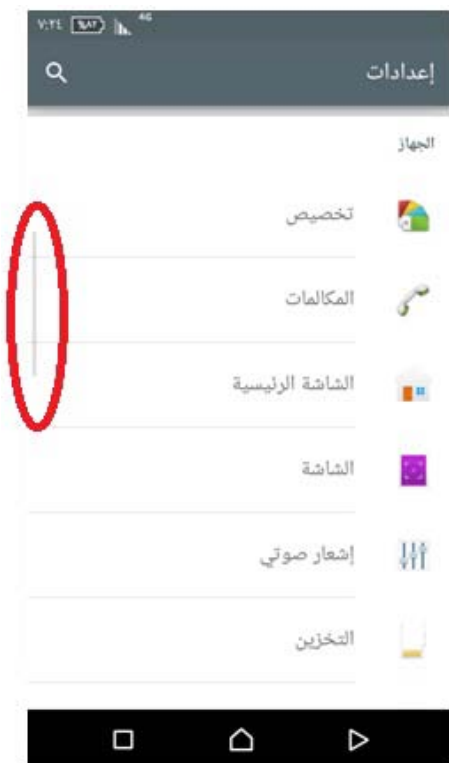


Figure 20. Example of a scroll bar positioned on the left side.

Text controls

Digit input

See Digits and Numbers in the Introduction as well as General reading/writing order in chapters above, but remember numbers are written from left-to-right in all BiDi languages. In a text field then, for instance a PIN input field or pass code field, the numbers shall be positioned as they are typed in and in the same way as in English (or any left-to-right-language). The only difference is that the numbers in the field should be **right-aligned**! See examples below:

Example - in English

1234

Example - in Arabic (localised numbers)

١٢٣٤

Example - in Arabic ("Western" numbers)



Text fields (cursor placement)

The default placement in any text field in a right-to-left language UI is to the very right – see example below. This placement is as mentioned a default and should look like below even before any character or text has been written into the field (please do not pay any attention to the screenshot below when it comes to things not translated).



Figure 21. Example of default placement of the cursor in a text field of a left-to-right language.

“Flipping through” a main menu, home screen

The so called pages in the main menu that you can go through by touching/“pushing” the screen to the left to go to the next page, shall be mirrored – this is a kind of progression and mirroring this goes hand in hand with any other progression, and should move from right to left..

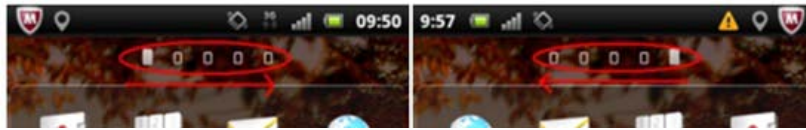


Figure 22. Example of flipping through screens in an English UI and a UI in a left-to-right language.

Tabs

Tabs shall be mirrored so that the top one is positioned to the very right and then the following ones fall under that and expand to the left. Look at the example below – these tabs are incorrect since they are still kept as they were in the left-to-right languages. All progression moves from right-to-left:

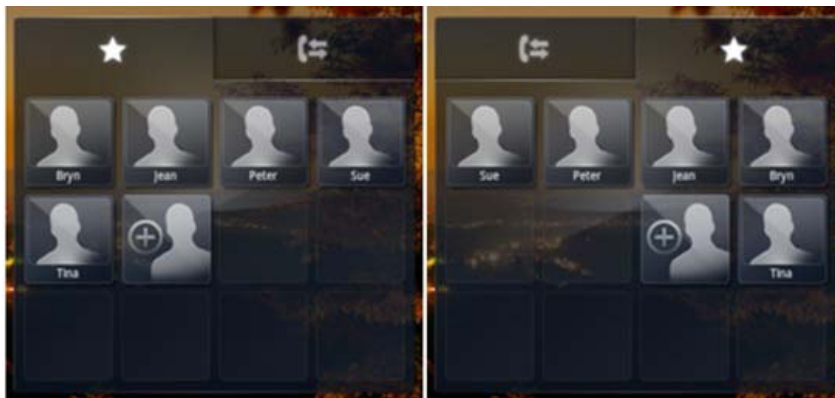


Figure 23. Example of a tab in English, and an incorrectly mirrored tab in a left-to-right language.

Date order

The numbers themselves are written left to right but the overall order is right-to-left.

Lock screen

Lock screen lock bar should not be mirrored

Status bar

UI should be mirrored, any progress bar should be mirrored, top status bar (including battery and clock) does not have to be mirrored but is preferable that it is.

Date & Time picker

- Date picker, UI should be mirrored.
- Time picker, Header should be mirrored and right-aligned, but not the time picker itself.

Theme picker

If there is any UI flow direction or flipping through, then it should be mirrored. Select buttons do not have to be mirrored but is preferable that they are.

Wallpaper picker

If there is any UI flow direction or flipping through, then the wallpaper should be mirrored. Select buttons do not have to be mirrored but is preferable that they are.

Sound picker

All UI should be mirrored, **however** Media player button arrow icon should be kept as is.

Music picker

All UI should be mirrored, **however** Player button arrow icon should be kept as is.

More information

For information regarding the latest Sony information, check out:

- <http://www.sony.com>