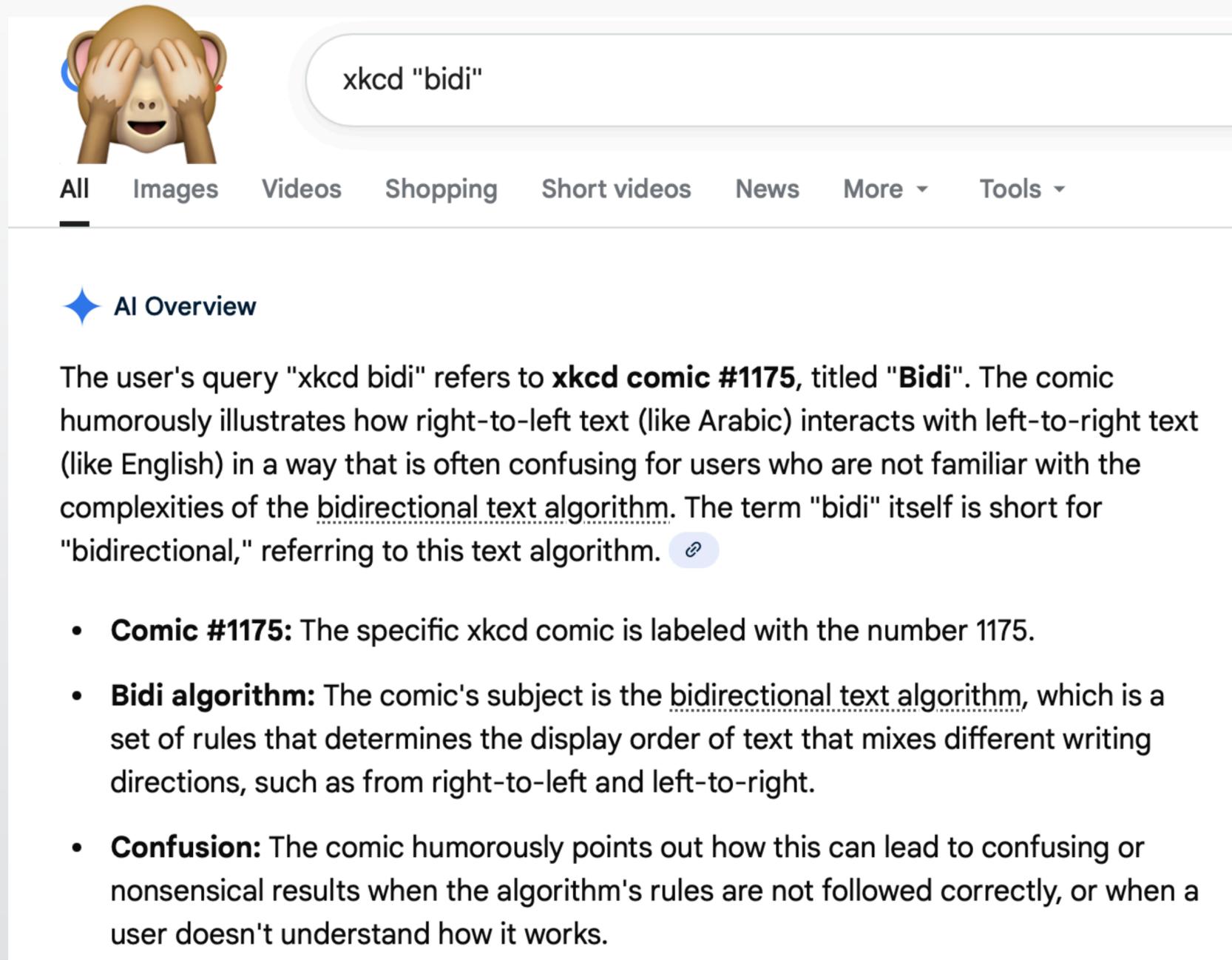
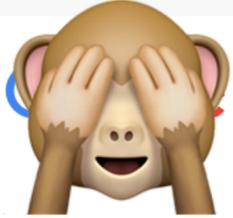


Ahmad Oustwan

# Bidi Text and UI (a tutorial)

# The bidi meme that never was



 xkcd "bidi"

All Images Videos Shopping Short videos News More Tools

◆ AI Overview

The user's query "xkcd bidi" refers to **xkcd comic #1175**, titled "**Bidi**". The comic humorously illustrates how right-to-left text (like Arabic) interacts with left-to-right text (like English) in a way that is often confusing for users who are not familiar with the complexities of the bidirectional text algorithm. The term "bidi" itself is short for "bidirectional," referring to this text algorithm. [🔗](#)

- **Comic #1175:** The specific xkcd comic is labeled with the number 1175.
- **Bidi algorithm:** The comic's subject is the bidirectional text algorithm, which is a set of rules that determines the display order of text that mixes different writing directions, such as from right-to-left and left-to-right.
- **Confusion:** The comic humorously points out how this can lead to confusing or nonsensical results when the algorithm's rules are not followed correctly, or when a user doesn't understand how it works.

# The bidi meme that was



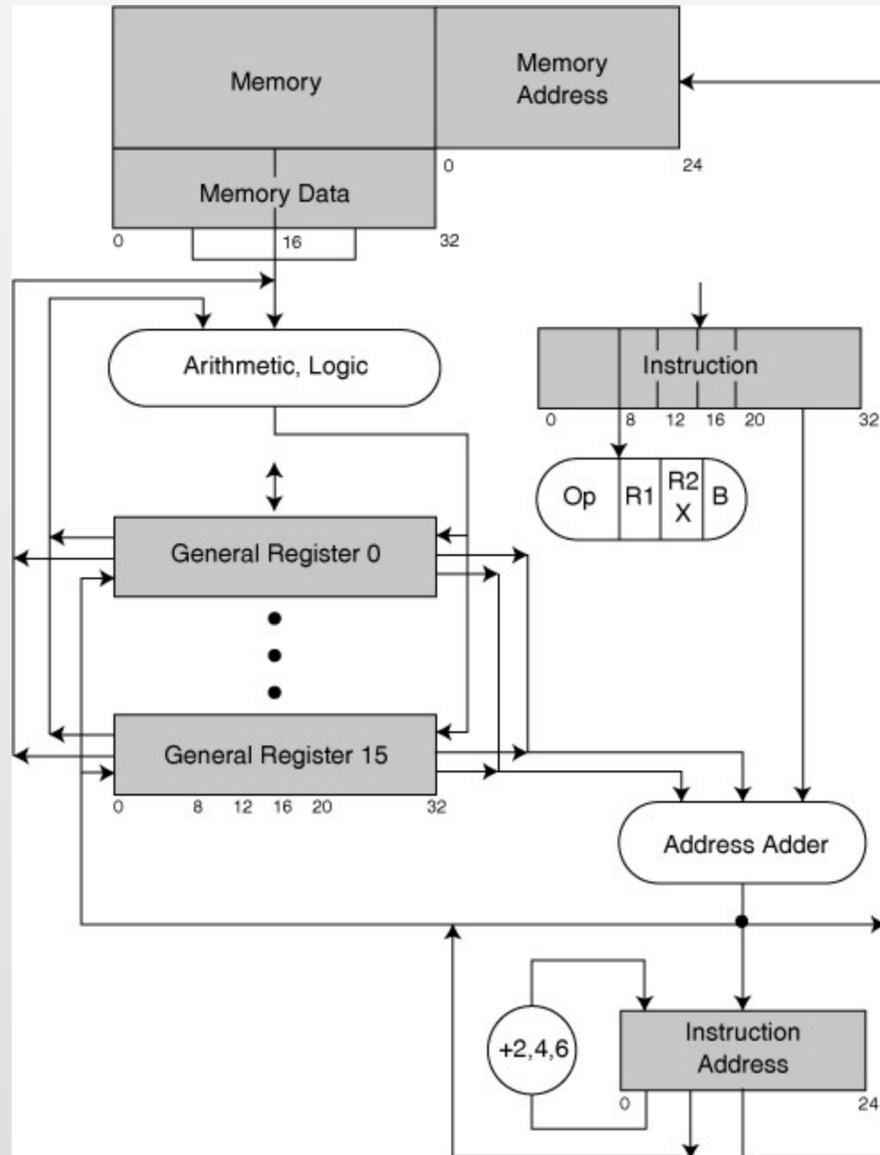
# A brief history of Bidi

# In the beginning

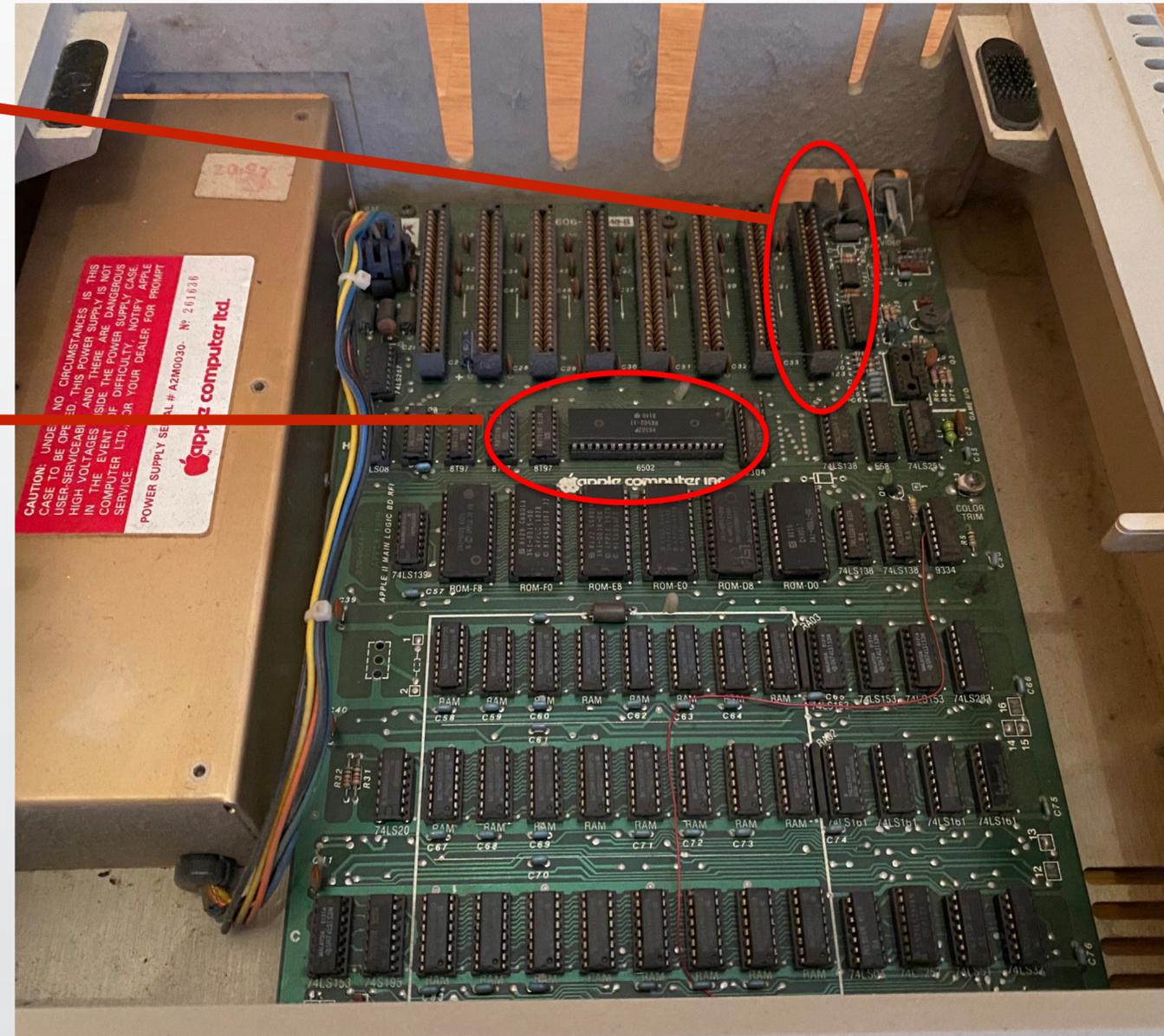


# In the beginning

Memory available for bilingual display code on extended display cards was only 500 bytes



6502  
Microprocessor



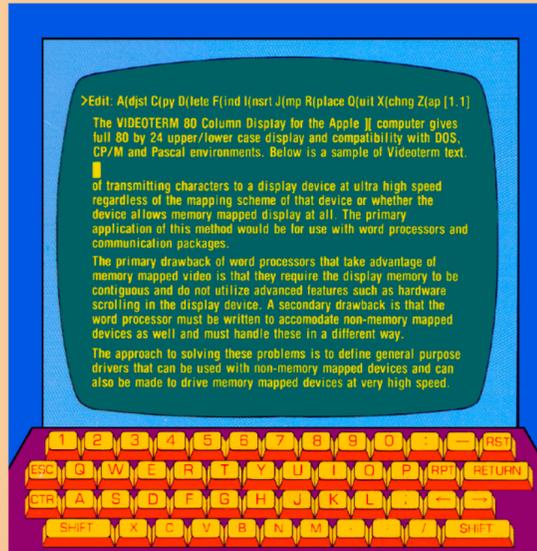
6502 Microprocessor architecture was modeled on the flexible IBM 360 mainframe, left. Roman-only screen display, Bios, was in a ROM Chip. control is the 6502 assembly instruction set, op-code.

# In the beginning



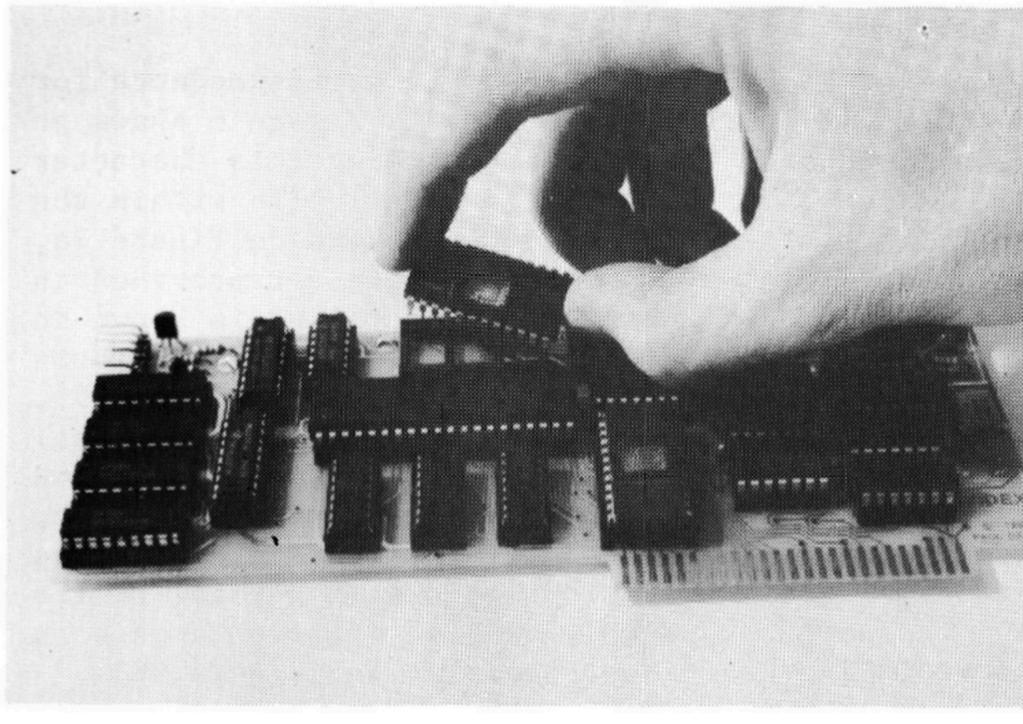
## Videoterm 80 Column Display

Installation and Operation Manual

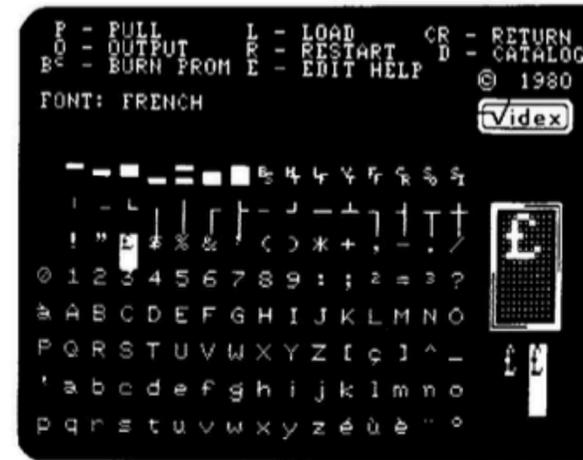


### A. Text

Naturally, you may print all 128 ASCII characters on your video monitor using the VIDEOTERM. However, the expanded character set available on the 2708 EPROM on the VIDEOTERM contains another 64 characters which may be printed.



### FRENCH:



### GERMAN:



### KATAKANA:



### SPANISH:







Apple Arabic Macintosh Plus,  
the power to succeed.



MF0204B



الكمبيوتر العربي "أبل ماكتوش بلاس"  
"السهل الممتنع"



MF0204B

# 7 bit to 8 bit to Unicode

ASMO449 > Mac Arabic > ISO 8859-6  
> Unicode

Moved to higher plane

Duplicated symbols but upper ones  
were strong RTL

8859-6 unified the symbols

That became Unicode and Unicode  
Bidi Algorithm

# 7 bit to 8 bit to Unicode

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0x	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
1x	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
2x	SP	!	"	#	¤	%	&	'	)	(	*	+	,	-	.	/
3x	0	1	2	3	4	5	6	7	8	9	:	;	>	=	<	?
4x	@	ء	آ	أ	ؤ	إ	ئ	ا	ب	ة	ت	ث	ج	ح	خ	د
5x	ذ	ر	ز	س	ش	ص	ض	ط	ظ	ع	غ	]	\	[	^	_
6x	-	ف	ق	ك	ل	م	ن	ه	و	ى	ي	´	¸		˘	¸
7x		¸	°									}		{	~	DEL

ASMO449 > Mac Arabic > ISO 8859-6  
> Unicode

Moved to higher plane

Duplicated symbols but upper ones  
were strong RTL

8859-6 unified the symbols

That became Unicode and Unicode  
Bidi Algorithm

# 7 bit to 8 bit to Unicode

ASMO 449 (1982)																
	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0x	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
1x	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
2x																
MacArabic <sup>[1]</sup>																
	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
2x	SP	!	"	#	\$	%	&	'	(	)	*	+	,./	-	./	/
3x	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4x	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5x	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6x	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7x	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL
8x	Ä	NBSP	Ç	É	Ñ	Ö	Ü	á	à	â	ä	ı	«	ç	é	è
9x	ê	ë	í	...	î	ï	ñ	ó	»	ô	ö	÷	ú	ù	û	ü
Ax	SP	!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
Bx	.	٠	١	٢	٣	٤	٥	٦	٧	٨	٩	:	:	<	=	>
Cx	*	ء	أ	أ	ؤ	إ	ئ	ا	ب	ة	ت	ث	ج	ح	خ	د
Dx	ذ	ر	ز	س	ش	ص	ض	ط	ظ	ع	غ	[	\	]	^	_
Ex	-	ف	ق	ك	ل	م	ن	ه	و	ى	ي	'	'	'	'	'
Fx	,	ء	°	پ	ٹ	چ	ه	ڦ	گ	ڻ	ڙ	{		}	ژ	ء

ASMO449 > Mac Arabic > ISO 8859-6 > Unicode

Moved to higher plane

Duplicated symbols but upper ones were strong RTL

8859-6 unified the symbols

That became Unicode and Unicode Bidi Algorithm

# 7 bit to 8 bit to Unicode

ASMO 449 (1982)																
	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0x	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
1x	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
2x																
MacArabic <sup>[1]</sup>																
	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
2x	SP	!	"	#	\$	%	&	'	(	)	*	+	,./	-	./	/
3x	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4x	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5x	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6x	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7x	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL
8x	Ä	NBSP	Ç	É	Ñ	Ö	Ü	á	à	â	ä	ı	«	ç	é	è
9x	ê	ë	í	...	î	ï	ñ	ó	»	ô	ö	÷	ú	ù	û	ü
Ax	S															
Bx	.															
Cx	*															
Dx	ن															
Ex	-															
Fx	○	○	○													

ASMO449 > Mac Arabic > ISO 8859-6 > Unicode

Moved to higher plane

Duplicated symbols but upper ones were strong RTL

8859-6 unified the symbols

That became Unicode and Unicode Bidi Algorithm

# Unicode Bidi Algorithm

- Arabic, Hebrew, etc are ~~RTL~~ Bi-Directional language
- Text is right-to-left  $D \leftarrow C \leftarrow B \leftarrow A$  ا ب ج د
- Numbers are left-to-right  $1 \rightarrow 2 \rightarrow 3 \rightarrow 4$  ١٢٣٤
- Symbols can be either  $٣٤..١٢$  ١٢.٣٤
- Special rules for math, numbers, punctuation

# What is the Bidi for ;-) ?

Dates must have control characters to format correctly.

**4/11/2017**

**112017/04/**

**2017/04/11**

Twitter: @name #subject

telephone: +1 (408) 234 5678 – but try rtl

URLs - [https://ar.wikipedia.org/wiki/الصفحة\\_الرئيسية](https://ar.wikipedia.org/wiki/الصفحة_الرئيسية)

Emoji .....

# Bidi Control Characters

- When UBA was wrong:
  - LRM (U+200E): Left-to-Right Mark
  - RLM (U+200F): Right-to-Left Mark
  - LRE (U+202A): Left-to-Right Embedding
  - RLE (U+202B): Right-to-Left Embedding
  - LRO (U+202D): Left-to-Right Override
  - RLO (U+202E): Right-to-Left Override
  - PDF (U+202C): Pop Directional Formatting

# Bidi Control Characters (in CLDR)

- Defined in CLDR to format numbers (see <https://st.unicode.org/cldr-apps/v#/ar/Symbols/>):

standard-currency	□#,##0.00  	□ ##0.00,#-;□ ##0.00,#  «RLM» #,##0.00 «NBSP» ⓘ □; «RLM» -#,##0.00 « NBSP» □	##0.00,# □ □ «NBSP» #,##0.00 ⓘ	
standard-currency-alphaNextToNumber	□ #,##0.00 □ «NBSP» ⓘ #,##0.00		□ ##0.00,#-;□ ##0.00,#  «RLM» #,##0.00 «NBSP» ⓘ □: «RLM» -#.##0.00 «	##0.00,# □ □ «NBSP» #,##0.00 ⓘ

- Intl.NumberFormat in Javascript (or similar) adds them

# But that was not enough

- What about inserting variables and names in apps?

عادل علاوي started playing stitch. 1mo ago

- Isolates added in 2010 to HTML/CSS, and 2012 to Unicode:
  - LRI (U+2066): Left-to-Right Isolate
  - RLI (U+2067): Right-to-Left Isolate
  - FSI (U+2068): First Strong Isolate
  - PDI (U+2069): Pop Directional Isolate

# Isolates in MessageFormat2

```
interface MessageFormatOptions {  
    bidiIsolation?: 'compatibility' | 'none';  
    dir?: 'ltr' | 'rtl' | 'auto';  
    functions?: { [key: string]: MessageFunction };  
    localeMatcher?: 'best fit' | 'lookup';  
}
```

# Higher level protocols



## three, four vs four, three

Punctuation takes a writing direction above the Unicode base direction

Let's try it

[ironymark.github.io/web-bidi/](https://ironymark.github.io/web-bidi/)

# Lifting the lid on Bidi



Natural Alignment

Visual and Logical Caret

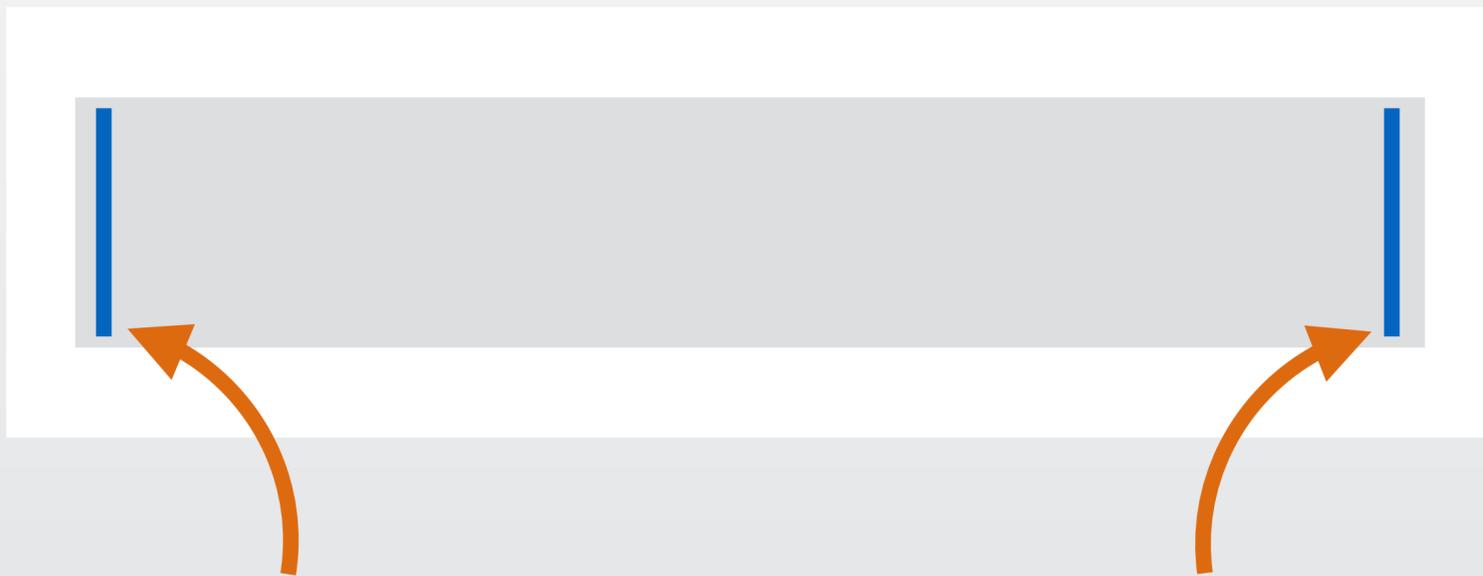
Visual and Normal Selections

Higher level protocols

RTL in tables

User Interface

# Natural alignment



What if the text box is empty

- Alignment is natural
- Keyboard/user context takes precedence
- Caret 'predicts' next character

# Natural alignment and direction

مدونة اللغة العربية

لم تأت حركة التطور العلمي والتقني الذي يتسم بها عصرنا الحاضر من فراغ، بل هي نتاج عملٍ دؤوب على مستويات مختلفة استمر قروناً عديدة.

التسلسل	المعجم
1	التهذيب
2	العياب
3	لسان العرب
5	القاموس المحيط
4	تاج العروس

Document available via: <https://www.egyptianlanguage.com/>

Paragraph style to:

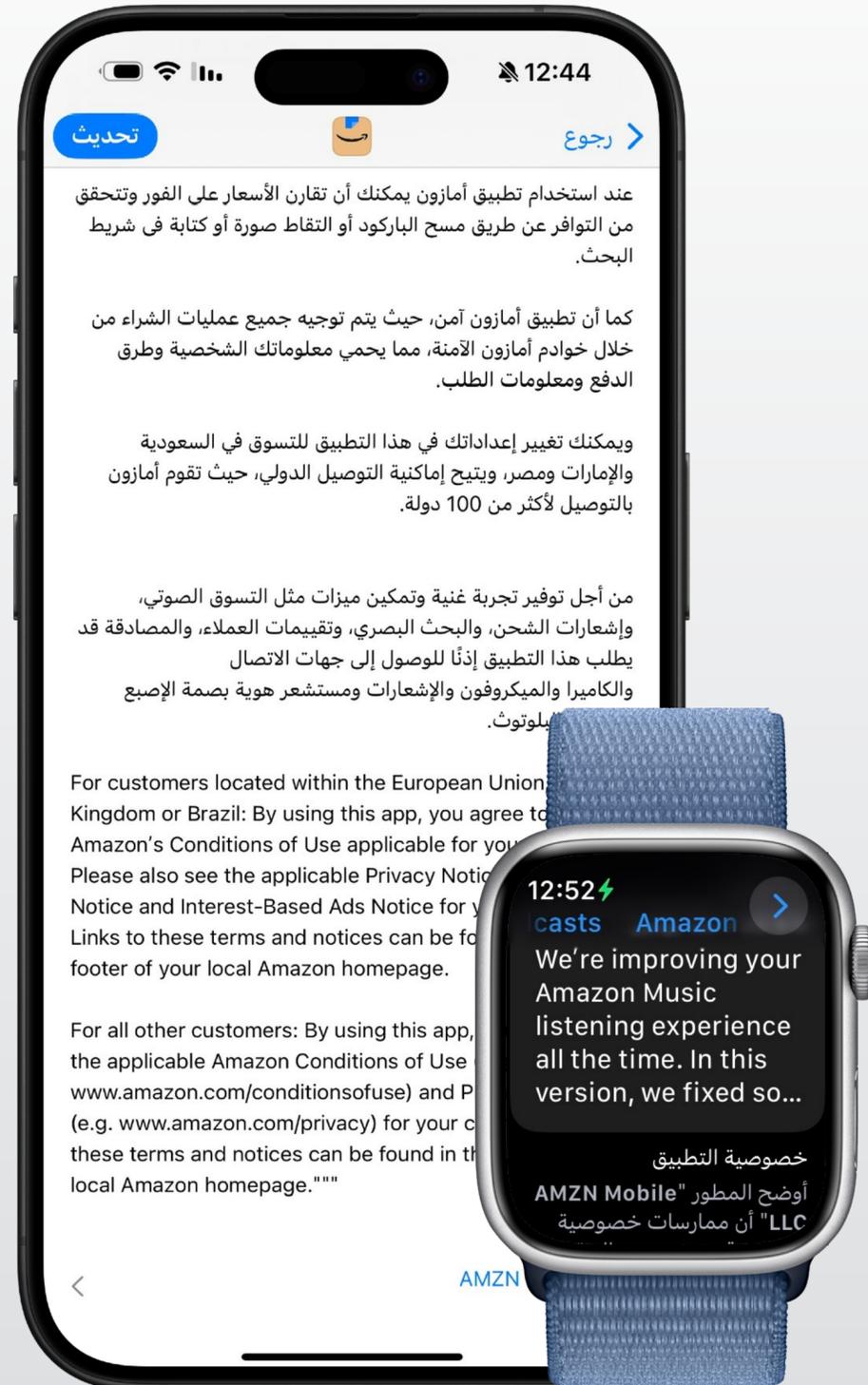
TextAlignment:natural

WritingDirection:natural

Container has:

Natural alignment

# Being Language Aware



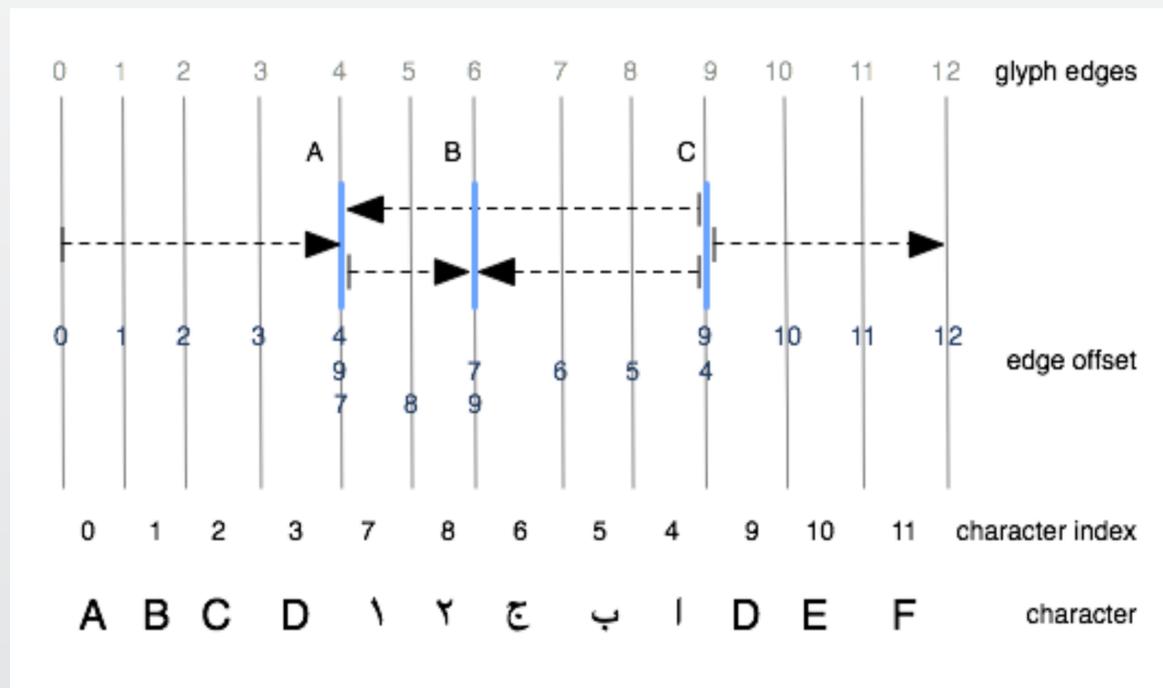
Text direction is important

Look at the content

# The Caret

عادل علاوي | Adil Allawi

# The Caret

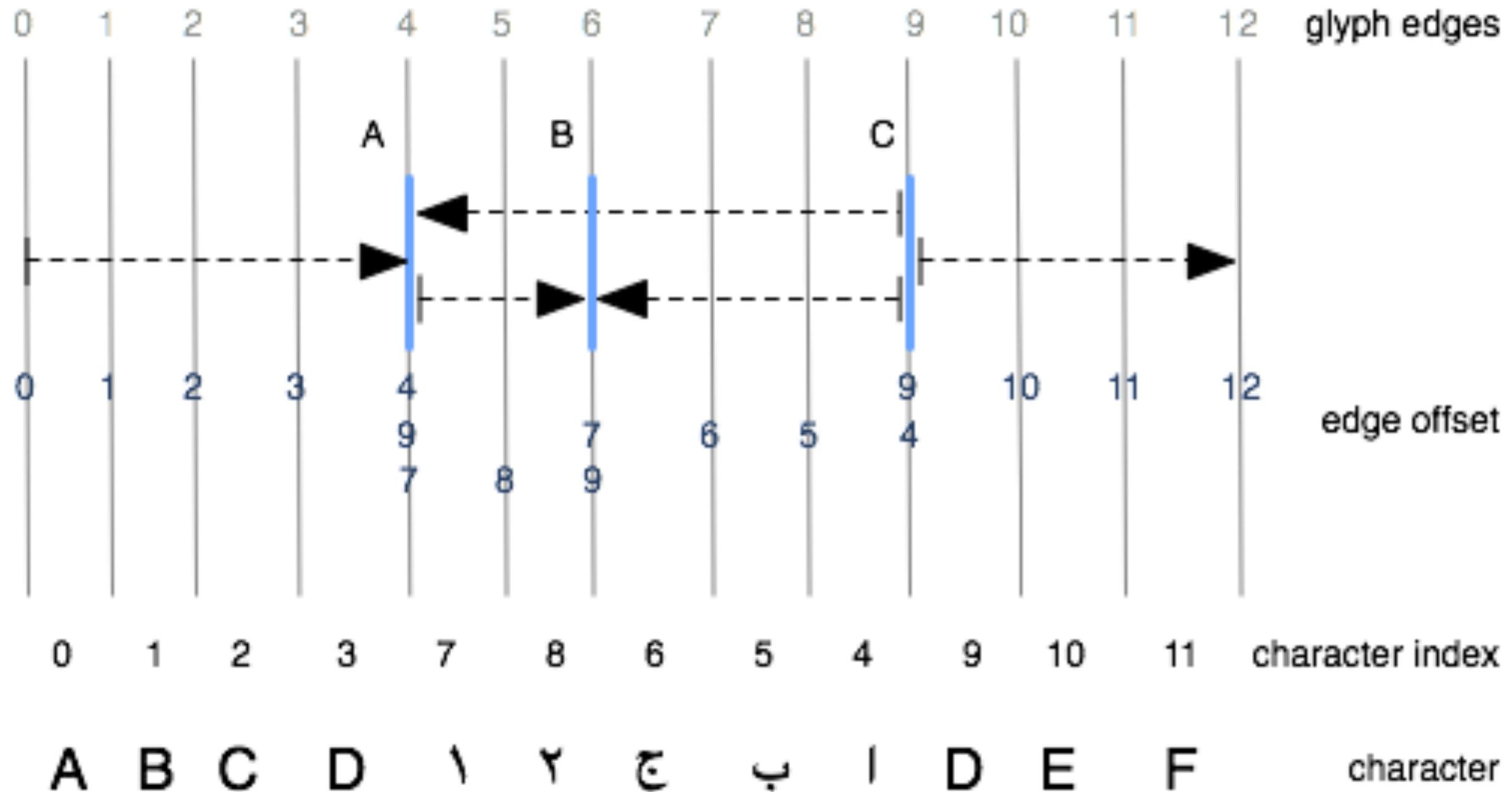


Single caret different locations

Caret has a character and an edge

Caret can represent multiple locations in the backing store

# The Caret



edge  
relations

# Natural vs Logical Caret

عادل علاوي | Adil Allawi

They look the same behave differently  
Caret changes according to user action

- Touch --> Visual
- Type --> Logical

# Logical Selection

علاوي عادل Adil Allawi

Complex outline

Simple range

# Natural Selection

A diagram illustrating text selection. The text "علاوي عادل Adil Allawi" is shown. A light blue rectangular selection box highlights the word "عادل" (Adil). A vertical blue line with circular endpoints at the top and bottom is positioned at the right edge of the selection box, indicating the selection direction. The text "علاوي" is in Arabic script, "عادل" is in Arabic script, "Adil" is in Latin script, and "Allawi" is in Latin script.

علاوي عادل Adil Allawi

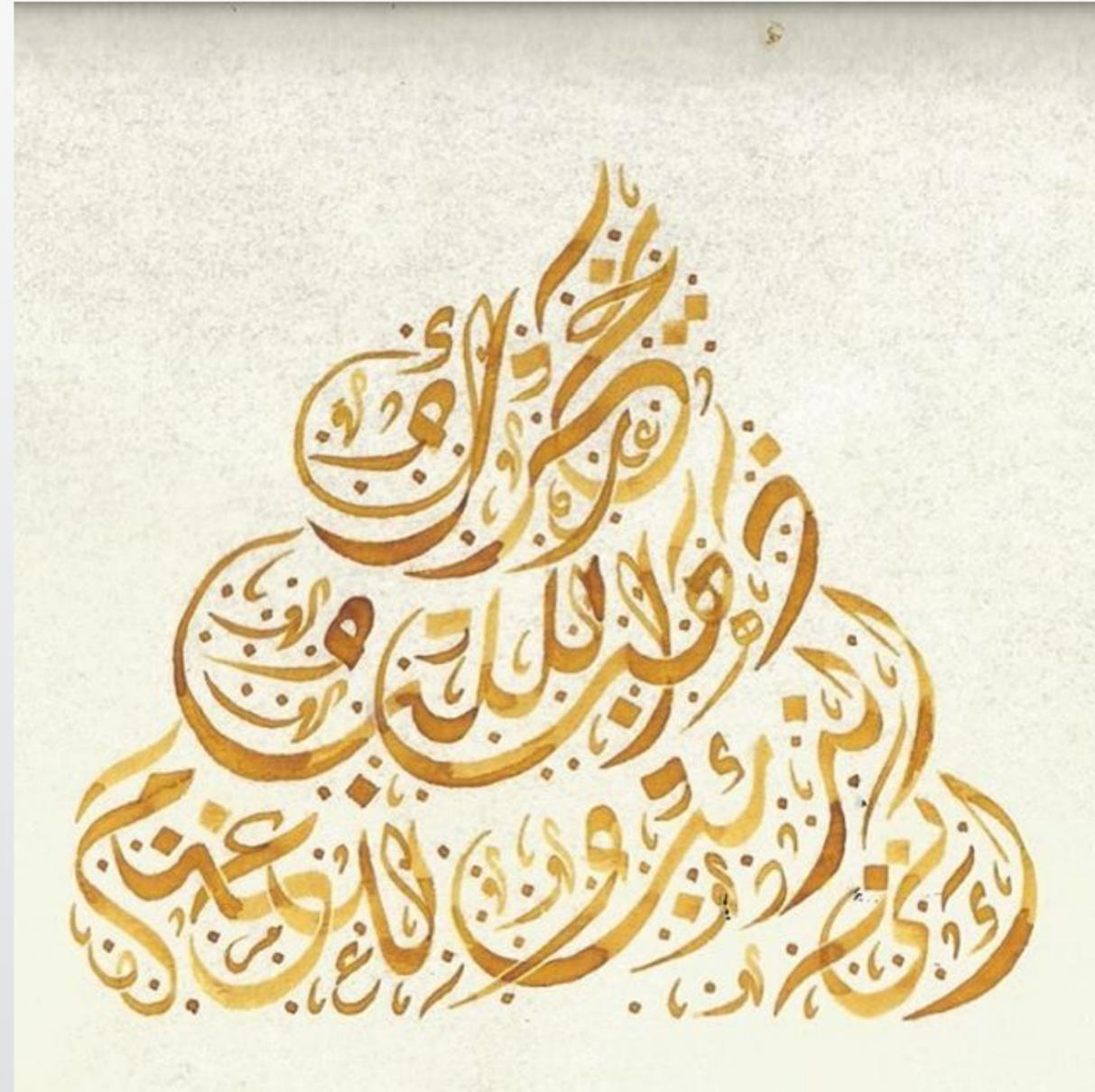
Select direction matches paragraph writing direction

Simple rectangle

Array of “visual” runs

Head and Tail

# UI Design



# Example

**الخط**

جيزة

عادي 11 نقطة

B I U

أنماط الأحرف لا شيء

**محاذاة**

← →

← →

↑ \* ↓

التباعد ١,٠ - مفرد

التعداد الرقمي والنقطي لا شيء

**Font**

Helvetica Neue

فاتح 48 pt

B I U

Character Styles None

**Alignment**

← →

← →

↑ \* ↓

Spacing 1.0 - Single

Bullets & Lists None\*

Inspectors from iWork

Note alignment and color picker are not reversed.

# RTL Tables

تم التعديل — بلا عنوان

تنظيم | التنسيق | مشاركة | تعليق | وسائط | شكل | نص | مخطط | جدول | إدراج | إضافة فئة | تكبير/تصغير | 125% | عرض

الورقة 1

الجدول 1

يناير ١٠, ١٩٩٩	يناير ٩, ١٩٩٩	يناير ٨, ١٩٩٩	يناير ٧, ١٩٩٩	١٩٩٩	يناير ٥, ١٩٩٩
يناير ١١, ١٩٩٩	يناير ١٠, ١٩٩٩	يناير ٩, ١٩٩٩	يناير ٨, ١٩٩٩	١٩٩٩	يناير ٦, ١٩٩٩
يناير ١٢, ١٩٩٩	يناير ١١, ١٩٩٩	يناير ١٠, ١٩٩٩	يناير ٩, ١٩٩٩	١٩٩٩	يناير ٧, ١٩٩٩
يناير ١٣, ١٩٩٩	يناير ١٢, ١٩٩٩	يناير ١١, ١٩٩٩	يناير ١٠, ١٩٩٩	١٩٩٩	يناير ٨, ١٩٩٩
يناير ١٤, ١٩٩٩	يناير ١٣, ١٩٩٩	يناير ١٢, ١٩٩٩	يناير ١١, ١٩٩٩	١٩٩٩	يناير ٩, ١٩٩٩
يناير ١٥, ١٩٩٩	يناير ١٤, ١٩٩٩	يناير ١٣, ١٩٩٩	يناير ١٢, ١٩٩٩	١٩٩٩	يناير ١٠, ١٩٩٩
يناير ١٦, ١٩٩٩	يناير ١٥, ١٩٩٩	يناير ١٤, ١٩٩٩	يناير ١٣, ١٩٩٩	١٩٩٩	يناير ١١, ١٩٩٩

Table 1-1

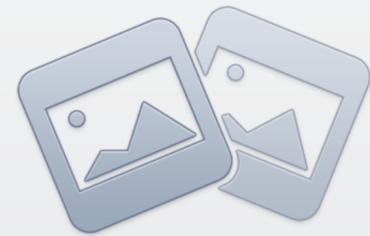
text		cell 3	cell 4	cell 5	cell 6
text		cell 4	cell 5	cell 6	cell 7
text		cell 5	cell 6	cell 7	cell 8
text		cell 6	cell 7	cell 8	cell 9
text		cell 7	cell 8	cell 9	cell 10
text		cell 8	cell 9	cell 10	cell 11
text		cell 9	cell 10	cell 11	cell 12

Tables should match text direction

Natural Direction is important

Cells can be numeric

# When to Mirror



Arrows for 'next' and 'previous'

Undo / Redo

Anything related to text

When an object is in layout direction



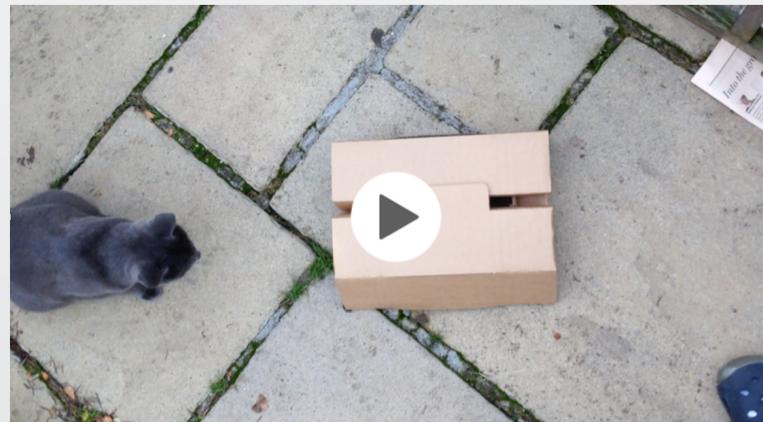
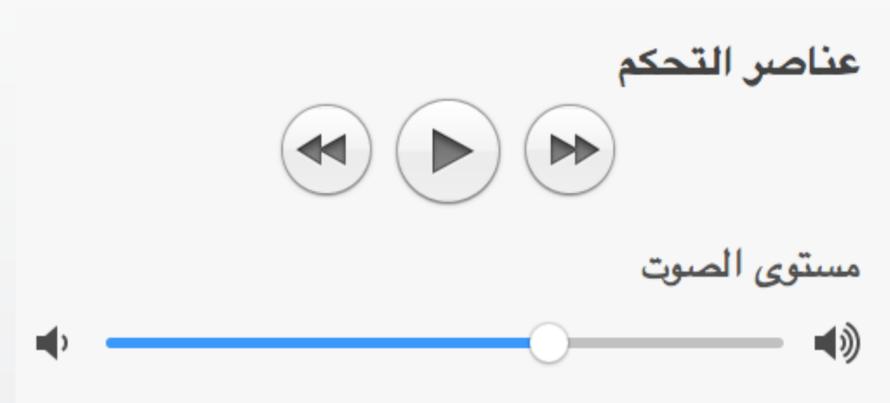
**Hyperlink**



**Hyperlink**



# When to Not Mirror



Rotation of time

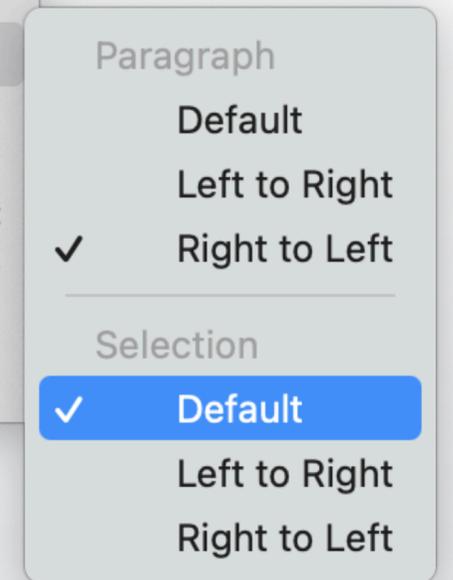
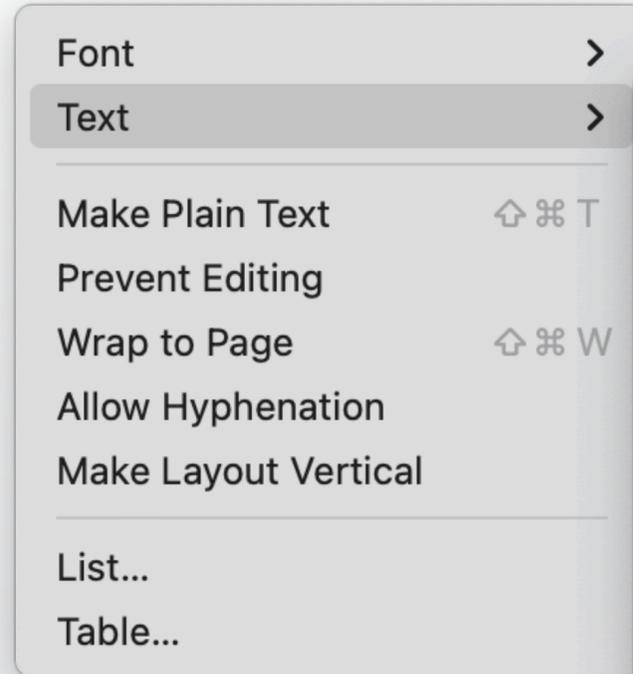
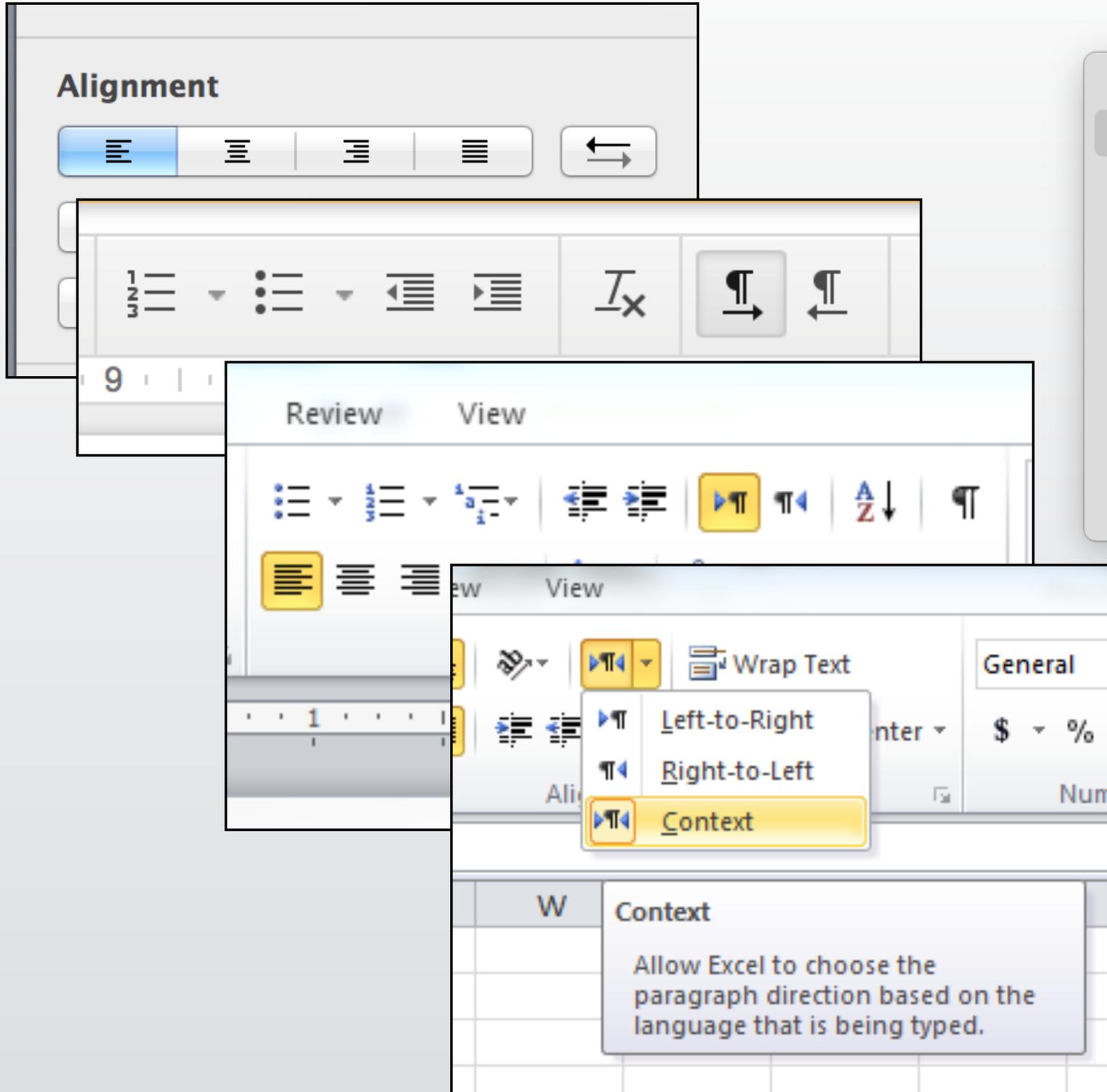
Physical objects

Movie controls

Image controls

Shadows

# Text Direction



# A fully RTL UI

Document Format Collaborate Comment Media Shape Text Chart Table Insert

79% Zoom View

Text Track Changes Highlight Comment 1 of 3 View Options

Update Fre...m\*

More Layout Style

Font

Muna

64 pt Bold

B I U

None Character Styles

Alignment

Spacing 1.0 - Single

None\* Bullets & Lists

Middleeastern\_cooking.pages

PAGE 1

Mar 28, 1:23 PM

Hello

Delete Reply

PAGE 4

Mar 28, 11:55 AM

This is a comment

Delete Reply

Mar 28, 11:55 AM

This is another comment

Delete Reply

أنواع الخطوط العربية

زاول أبو حيان صناعة الوراقة ونسخ الكتب

العربية التي كان منها ما هو مستعمل قديماً ، ومنها قريبة الحديث ، أما هذه الطرائق المستنبطة فهي مروية عن الصحابة حتى اتصلت بابن مقلة وإفوت وغيرهم . وهم تفننوا فيها حسب اجتهادهم . وكنسأ - أطال الله بقاءك - في مجلس ابن البربري وقد حفل بإرساب الأقسام والخطوط ، وصار كل منهم يظهر منيأته من النوادر .

أنواع الأقسام : فقال احدهم : خير الكلام ما استمكن نُصجه في جرمه ، وجف ماؤه في قشره ، وقطع

زاول أبو حيان صناعة الوراقة ونسخ الكتب قائم بإصولها وقواعدها واطلع على دقائقها واصنافها وها هو ذا يقول فيها كلمة العالم الخبير : أنواع الخطوط العربية : الإسماعيلي ، والمكي ، والمدني ، والأندلسي ، والشامي ، والسعراقي ، والعباسي ، والشعب ، والرئحاني ، والمجرد ، والمصري ، فهذه هي الخطوط

Start from the user and  
work backwards