

Cultural assumptions and localization tailoring: The dilemma of numbers bi-directionality in the Arabic script

Mokhtar BEN HENDA

ISD, Manouba University, Tunisia

Mokhtar.Benhenda@isd.run.tn

Context

- ❑ An experience from the «South» where E-Learning systems and education markets are having a growing role
- ❑ Tunisia : A Virtual University with an objective to put On-line 20% of educational contents by 2006
- ❑ This exposé : some outcomes of experiments and fundamental research aimed at defining the most appropriate localization solutions for multilingual and multicultural educational contents within an Arabic educational context

Localization can not take place without a strict implication of Internationalization within the frame of a Globalization context

Definitions

Localisation (L10n)

- ❑ Localization is the process of preparing a product to cope with the requirements of a target market different from that for which it was originally developed.
- ❑ It implies not only a mere translation of the user interface, but also an application of linguistic and cultural parameters in respect with local regulations and standards

WEM, Lisbon 20-22 May 2003 - Cultural Assumptions and L10n Tailoring - The Arabic Script, Mokhtar BEN HENDA, ISD, Manouba University, TUNISIA

Localization can not take place without a strict implication of Internationalization within the frame of a Globalization context

Definitions

Internationalization (I18n)

- ❑ Internationalization is the process of developing a piece of software from the ground up so that it can be translated for a new user community without recompilation.
- ❑ An internationalized application:
 - Displays messages, data, text, etc. according to local conventions
 - Doesn't need recompilation to run in a new location

WEM, Lisbon 20-22 May 2003 - Cultural Assumptions and L10n Tailoring - The Arabic Script, Mokhtar BEN HENDA, ISD, Manouba University, TUNISIA

Localization can not take place without a strict implication of Internationalization within the frame of a Globalization context

Definitions

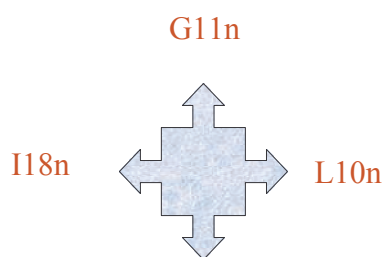
Globalization (G11n)

- ❑ Generated simultaneously by the application of new technologies, which overcome barriers of time and space, and the generalization of the principles of free trade in a market of world dimensions (Global Market).
- ❑ In itself, this phenomenon does not necessarily imply the homogenisation of contents which threatens cultural diversity.

WEM, Lisbon 20-22 May 2003 - Cultural Assumptions and L10n Tailoring - The Arabic Script, Mokhtar BEN HENDA, ISD, Manouba University, TUNISIA

Localization can not take place without a strict implication of Internationalization within the frame of a Globalization context

Definitions



Cultural and linguistic issues to face

WEM, Lisbon 20-22 May 2003 - Cultural Assumptions and L10n Tailoring - The Arabic Script, Mokhtar BEN HENDA, ISD, Manouba University, TUNISIA

The Challenges

1 - Deal with the the **Multilingualism** of «complex» scripts

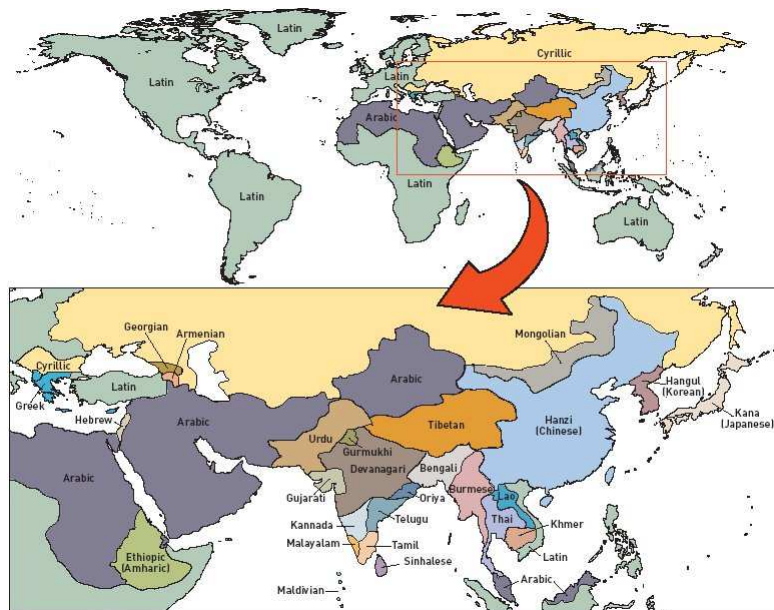
- Characters rendering (fonts and glyphs)
- Bi-Directionality (RTL : LTR)

2 - Deal with the the **cultural diversity** of target markets

- Data systems for social regulation (Calendar, Time)
- Indigenous cultural inheritance (Psycho-Cognitive matters)

WEM, Lisbon 20-22 May 2003 - Cultural Assumptions and L10n Tailoring : The Arabic Script, Mokhtar BEN HENDA, ISD, Manouba University, TUNISIA

The World Writing Systems Geographic Distribution



WEM, Lisbon 20-22 May 2003 - Cultural Assumptions and L10n Tailoring : The Arabic Script, Mokhtar BEN HENDA, ISD, Manouba University, TUNISIA

«Complex» Writing Systems Geographic Distribution

Language	Number of Countries
Arabic	18
Farsi	1 (Iran)
Urdu	2 (India, Pakistan)
Hebrew	1 (Israel)
Yiddish	Israel, North America, South America, Russia, Europe

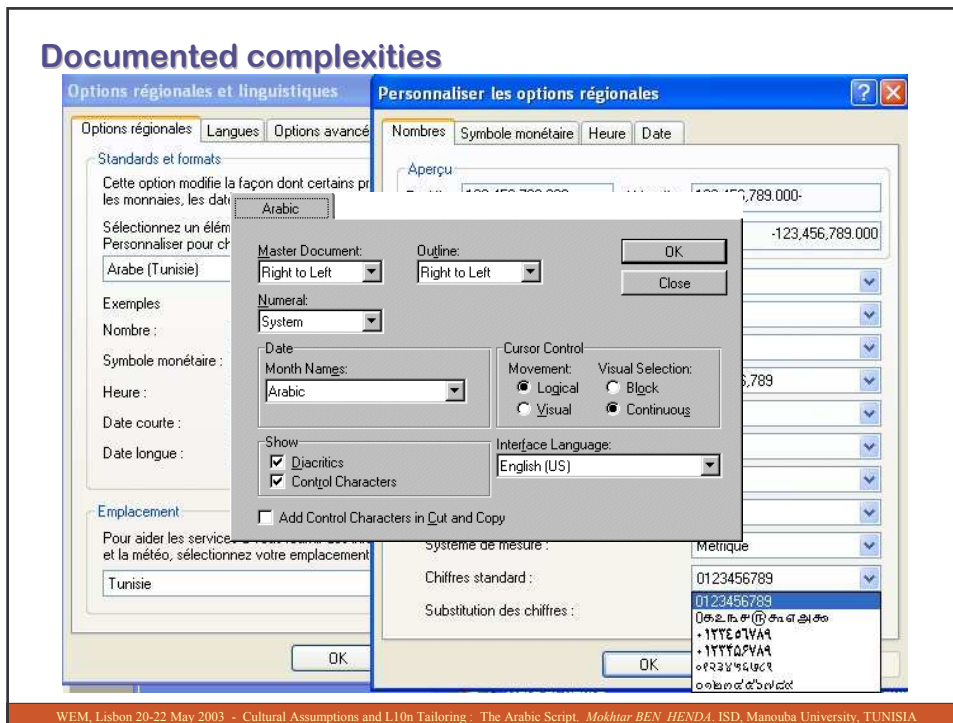
WEM, Lisbon 20-22 May 2003 - Cultural Assumptions and L10n Tailoring : The Arabic Script, Mokhtar BEN HENDA, ISD, Manouba University, TUNISIA

Documented complexities

- ❑ Most cultural and linguistic parameters are documented and treated through Locales
- ❑ A locale is a piece of software that allows determining a set of variables that identify a specific region for which Internationalization and Localization processes are supposed to be done. Locale variables may specify :
 - Bi-Directional Writing systems
 - Character Sets and glyph rendering
 - Numeric systems (*Hindu, Arabic, Roman etc.*)
 - Calendar systems (*Gregorian, Julian, Islamic, Chinese etc.*)
 - Measuring units (*weights, distance*)
 - Hour / Date formats
 - Currency symbols
 - Time zone
 - Geographic location
 - ...

WEM, Lisbon 20-22 May 2003 - Cultural Assumptions and L10n Tailoring : The Arabic Script, Mokhtar BEN HENDA, ISD, Manouba University, TUNISIA

Documented complexities



WEM, Lisbon 20-22 May 2003 - Cultural Assumptions and L10n Tailoring : The Arabic Script, Mokhtar BEN HENDA, ISD, Manouba University, TUNISIA

Undocumented complexities

The case of the Arabic language

Bi-directionality of Numbers and Cultural assumptions

1st Pre-requisite :

The arithmetic incremental algorithm used in the decimal counting system is a RTL process

Thousands	Hundreds	Tens	Units
0	0	0	1
0	0	0	9
0	0	1	0
0	0	9	9
0	1	0	0
0	9	9	9
1	0	0	0
~~~~~			
1	9	8	7
1	9	8	8
1	9	8	9
1	9	9	0
~~~~~			
1	9	9	8
1	9	9	9
2	0	0	0

WEM, Lisbon 20-22 May 2003 - Cultural Assumptions and L10n Tailoring : The Arabic Script, Mokhtar BEN HENDA, ISD, Manouba University, TUNISIA

Undocumented complexities

1st Pre-requisite :

- Oriental Origins of the decimal counting system
- Copes with Standard Arabic writing system
(Handwriting, Counting, Pronunciation)

ARABE STANDARD

Transcription

✍ ... 999.999 - 99.999 - 9.999 - 999 - 99

Valeurs à positions

✋ ... 999.999 - 99.999 - 9.999 - 999 - 99

يُعتبر الرقم 1.256.432 أو القيمة 27% مطابقان لفضاً أو نسخاً
لاتجاه الكتابة العربيّة من اليمين إلى اليسار في إطار لغويّ مُقنّن.

WEM, Lisbon 20-22 May 2003 - Cultural Assumptions and L10n Tailoring - The Arabic Script, Mokhtar BEN HENDA, ISD, Manouba University, TUNISIA

Undocumented complexities

The case of the Arabic language

Bi-directionality of Numbers and Cultural assumptions

2nd Pre-requisite:

Imported to Occident, this system received a second functional layer to adapt it to the Latin LTR writing system

⇒ *A first bias*

✍ ≠ ✋

Le même chiffre 1.256.235 ou le taux 27 %

accusent une opposition entre le sens de la transcription alphabétique et l'incrémentation quantitative des chiffres

Saisie →

3
6
9

Calculabilité ↓

✋ 99 - 999 - 9

3	789
4	1368
5	

= 1 3 6 8

← Valeur à position

WEM, Lisbon 20-22 May 2003 - Cultural Assumptions and L10n Tailoring - The Arabic Script, Mokhtar BEN HENDA, ISD, Manouba University, TUNISIA

Undocumented complexities

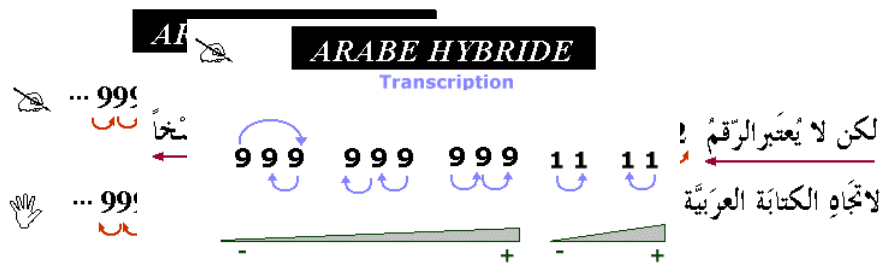
The case of the Arabic language

Bi-directionality of Numbers and Cultural assumptions

3rd Pre-requisite:

The new modification was transported back to the Arab world
(Industrial revolution / colonial period)

⇒ *A Second (doubled) bias*



WEM, Lisbon 20-22 May 2003 - Cultural Assumptions and L10n Tailoring - The Arabic Script, Mokhtar BEN HENDA, ISD, Manouba University, TUNISIA

Undocumented complexities

The case of the Arabic language

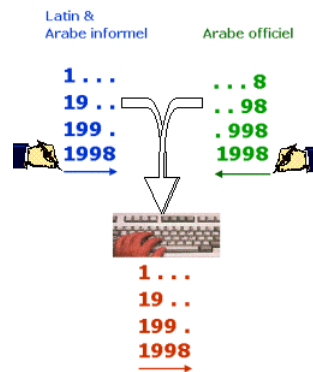
Bi-directionality of Numbers and Cultural assumptions

4th Pre-requisite:

Automated systems force a LTR number orientation

⇒ *A Third bias*

A cultural conflict and two different references for official Arabic users (cultural and technological).



WEM, Lisbon 20-22 May 2003 - Cultural Assumptions and L10n Tailoring - The Arabic Script, Mokhtar BEN HENDA, ISD, Manouba University, TUNISIA

Undocumented complexities

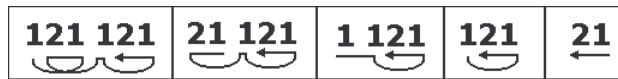
The case of the Arabic language

Bi-directionality of Numbers and Cultural assumptions

5th Pre-requisite:

The oral structure of numbers reinforces the bias

⇒ A Fourth bias



WEM, Lisbon 20-22 May 2003 - Cultural Assumptions and L10n Tailoring - The Arabic Script, Mokhtar BEN HENDA, ISD, Manouba University, TUNISIA

Undocumented complexities

The case of the Arabic language

Bi-directionality of Numbers and Cultural assumptions

The state of art :

- Two prevailing number systems : formal and informal (for oral and written modes)
- I18n and L10n Locales do not correspond neither to the first nor to the second mechanism
- Although the informal Arabic counting system is new, it has its sociolinguistic basis (*It has the same structure as the German system !?*)

WEM, Lisbon 20-22 May 2003 - Cultural Assumptions and L10n Tailoring - The Arabic Script, Mokhtar BEN HENDA, ISD, Manouba University, TUNISIA

Undocumented complexities

The case of the Arabic language

Bi-directionality of Numbers and Cultural assumptions

Num. Values	German	Arabic ^(*)	French	English
1	Eins	Un	Un	One
10	Zehn	Dix	Dix	Ten
11	Elf	Un et Dix	Onze	Eleven
12	Zwölf	Deux et Dix	Douze	Twelve
19	Neunzehn	Neuf et Dix	Dix Neuf	Nineteen
20	Zwanzig	Vingt	Vingt	Twenty
21	Eins Und Zwanzig	Un et Vingt	Vingt et Un	Twenty One
99	Neun Und Zwanzig	Neuf et Quatre Vingt Dix	Quatre Vingt Dix Neuf	Ninety nine
100	Ein Hundert	Cent	Cent	One hundred
101	Ein Hundert Eins	Cent et Un	Cent et Un	One Hundred One
121	Ein Hundert Eins Und Zwanzig	Cent et Un et Vingt	Cent Vingt et Un	One hundred Twenty One
1998	Ein Tausend neun Hundert Acht Und Neunzig	Mille et Neuf Cent et Huit et Quatre Vingt Dix	Mille Neuf cent Quatre Vingt Dix Neuf	One Thousand Nine Hundred Ninety Eight

(*) = Hybrid Arabic transliterated for communicability reasons and comparison with Latin equivalents

Comparisons between morphologic structures of numbers in different languages

WEM, Lisbon 20-22 May 2003 - Cultural Assumptions and L10n Tailoring - The Arabic Script, Mokhtar BEN HENDA, ISD, Manouba University, TUNISIA

And education in all that ?

- We led an experience on students using bilingual interfaces and discovered that these issues are not transparent to them. They slow the process of data processing and knowledge acquisition
- We already suggested to take these notes into consideration for the localization of coming versions of LOM elementary scheme (Learning Object Metadata) launched by ISO-SC36 to describe E-Learning contents (9 categories and tens of items)

WEM, Lisbon 20-22 May 2003 - Cultural Assumptions and L10n Tailoring - The Arabic Script, Mokhtar BEN HENDA, ISD, Manouba University, TUNISIA

Conclusion : *recommandations*

Considering that ...

- When you produce computer software, two thirds of your potential market is outside the English-speaking world
- People expect your product to “just work”
- Many users cannot or will not use a program that does not interact with them in their infinite and subtle language and cultural references
- A program is useless to a user when it can't handle the data he needs in the way he expects

Then you have to ...

- Localize relying on cultural and linguistic assets provided by people living in the target market more than experts having knowledge about it (even if they are native speakers but living outside the target market)

WEM, Lisbon 20-22 May 2003 - Cultural Assumptions and L10n Tailoring : The Arabic Script, Mokhtar BEN HENDA, ISD, Manouba University, TUNISIA

Thank You

WEM, Lisbon 20-22 May 2003 - Cultural Assumptions and L10n Tailoring : The Arabic Script, Mokhtar BEN HENDA, ISD, Manouba University, TUNISIA