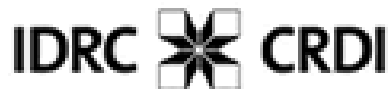


An introduction to localisation in Africa

"Regional Consultation on Local Language
Computing Policy in Developing Asia"
22-24 January 2007, Thimphu, Bhutan





PanAfrican Localisation Project

- Consists of 3 main parts:
 - Survey Document; Workshop; Resource website (growing wiki; 3 email lists)
- Need to better understand localisation in Africa – actual & potential – in order to move forward
- Need to assist localisers, decision makers, etc. with:
 - Networking (mutual aid, harmonization)
 - Resources (information, training, more?)
- IDRC, Kabissa, Bisharat

Where to start?

- 53 countries + other territories
- Many indigenous languages
 - *Ethnologue* counts over 2000
- Legacy of colonial history
 - European languages
 - Language communities divided
- Multiple realities
 - Differences among regions
 - Offset of languages & countries

Goals for this talk

- Describe the context of localisation
 - ICT overview
 - Describing the language situation
- Describe localisation in Africa
- Describe *PanAfrican Localisation* project
- Keys to success & future prospects

Definitions

- “African languages”
 - 4 main language families in Africa plus indigenised languages
- “Localisation”
 - Input methods (& last mile of i18n)
 - Software
 - Content (including local content creation)
- “Policy”
 - Public (govt.), but also donor & NGO

Basic constraints to L10n in Africa (esp. W, E, S)

- Technical
 - Low / unreliable electrification
 - Low numbers of computers
 - Older systems
 - Low / poor internet connectivity
- Educational
 - Low literacy / pluriliteracy
- Language policy
 - Low priority to indigenous languages
- Other language-related issues
 - Sociolinguistics, demographics, attitudes

ICT “environmental” factors affecting L10n

- Expanding country connectivity
- Country ICT policies (per UNECA)
 - Little if any mention of localisation
- Various development agencies initiating ICT projects, small & large
- Community radio is important
- Rapid expansion of use of cellphones
 - In many cases subscribership is far above fixed line use (& availability)

Africa & Asia internet comparison

- African internet use rates are much less than in Asia

– Source: <http://www.internetworldstats.com>

	Population (2007 Est.)	Pop. % in World	Internet Users, Latest Data	Penetration (% Pop.)	% Users in World	Use Growth (2000-07)
Africa	933,448,292	14.2 %	32,765,700	3.5 %	3.0 %	625.8 %
Asia	3,712,527,624	56.5 %	389,392,288	10.5 %	35.6 %	240.7 %
WORLD	6,574,666,417	100.0 %	1,093,529,692	16.6 %	100.0 %	202.9 %

Language in Africa: Differences from Asia

- African language communities
 - usually cross borders
 - are smaller on average than those in Asia
- Fewer writing systems in Africa
 - More use of Latin alphabet, often modified
 - Arabic, Ethiopic/Ge'ez, Tifinagh, N'Ko
- Shorter written tradition mostly; strong oral traditions (but increasing writing)
- Greater reliance on former colonial languages for government & education
- Ambivalent attitudes to African langs

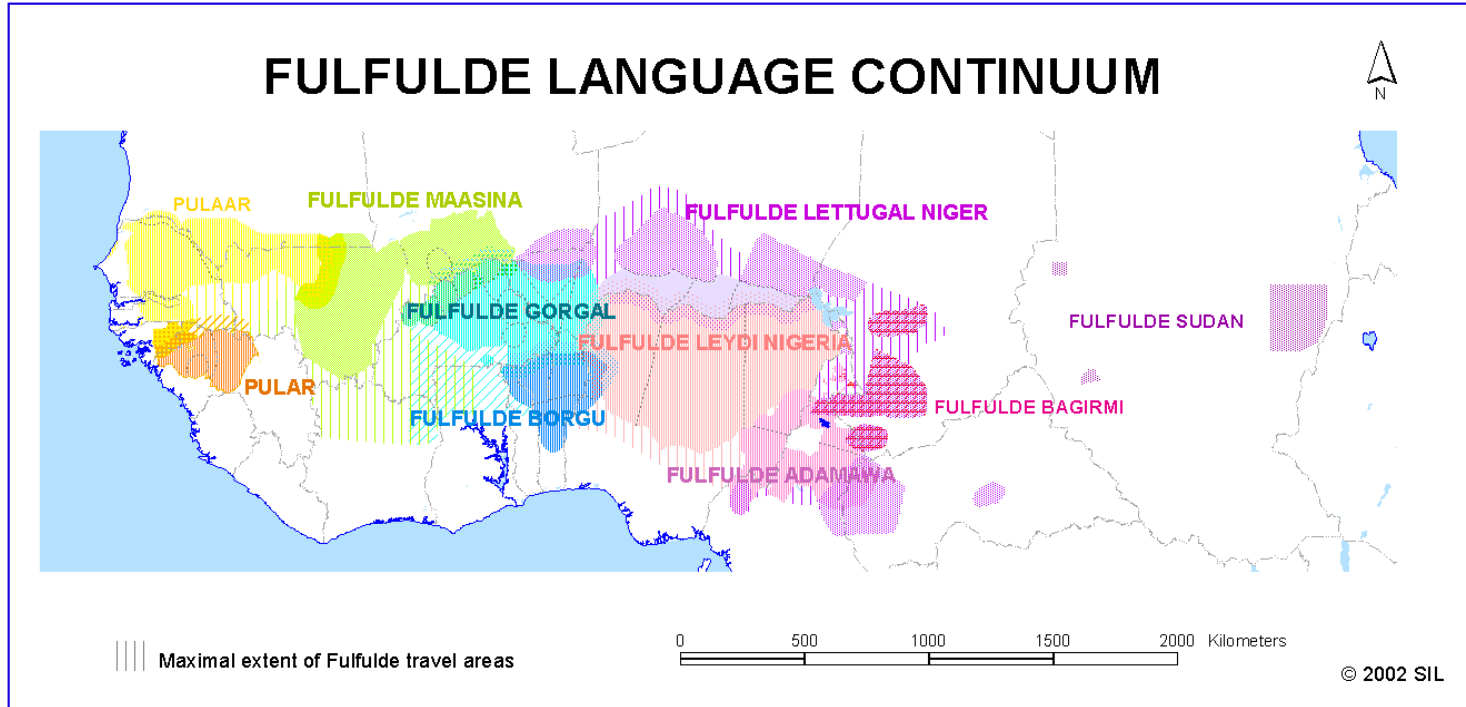
Africa & Asia language speakers comparison

- Using *Ethnologue's* counts, Africa has a lot more “languages” per population
- With continental population less than the country of India, it has almost 5x the number of “languages”

	Living languages		Number of speakers			
	<i>Count</i>	<i>Percent</i>	<i>Count</i>	<i>Percent</i>	<i>Mean</i>	<i>Median</i>
AFRICA	2,092	30.3 %	675,887,158	11.8 %	323,082	25,391
ASIA	2,269	32.8 %	3,489,897,147	61.0 %	1,538,077	10,171
<i>India</i>	428		1,065,070,607		2,448,483	
WORLD	6,912	100.0 %	5,723,861,210	100 %	828,105	7,000

Complex distribution: Example of Fula

- 8-11 varieties, no single standard
- 20m speakers in 15 countries
- Overlap with other languages all over



Writing systems (1/6)

- Several scripts used in Africa, though the Latin script dominates, often modified and/or with diacritics
 - Unicode has helped a lot for the latter
 - Latin A & B, IPA, Extended Latin additional

D

d
U0257

Ō
U1ECC

ō
U1ECD

Writing systems (2/6)

- Arabic script for Arabic, of course
- Also used historically and to some degree today for some other African languages (as "Ajami")
 - An example from Hausa (not standardized):

b	ب	g	گ	m	م	ts	ټ
β	ب	h	ه	n	ن	w	و
c	س	j	ج	r	ر	y	ي
d	د	k	ك	s	س	z	ز
ɗ	ط	ƙ	ك	sh	ش	,	,
f	ف	l	ل	t	ت		

Writing systems (3/6)

- Ethiopic/Ge'ez syllabary

- Used mainly for Amharic and Tigrinya

- Ex:

	hoy	lāwe	hāwt	may	śāwt	ra's	sat	š	qaf	qaf									
ሀ	hā	ለ	lā	ሐ	hā	መ	mā	ሠ	śā	ረ	rā	ሰ	sā	ሸ	šā	ቀ	qā	ቁ	quā
ሁ	hu	ሉ	lu	ሐ	hu	ሙ	mu	ሠ	śu	ሩ	ru	ሰ	su	ሸ	šu	ቀ	qu		
ሂ	hi	ሊ	li	ሐ	hi	ሚ	mi	ሢ	śi	ሪ	ri	ሲ	si	ሸ	ši	ቂ	qi	ቃ	qui
ሃ	ha	ላ	la	ሐ	ha	ማ	ma	ሣ	śa	ራ	ra	ሳ	sa	ሸ	ša	ቃ	qa	ቄ	qua
ሄ	he	ሌ	le	ሐ	he	ሜ	me	ሣ	śe	ራ	re	ሴ	se	ሸ	še	ቄ	qe	ቅ	que
ሀ	ha/ø	ለ	la/ø	ሐ	ha/ø	ሞ	ma/ø	ሠ	śa/ø	ረ	ra/ø	ሰ	sa/ø	ሸ	ša/ø	ቀ	qa/ø	ቃ	qua
ሀ	ho	ሎ	lo	ሐ	ho	ሞ	mo	ሠ	śo	ረ	ro	ሰ	so	ሸ	šo	ቀ	qo		
	bet	tāwe	č		hārm	nāhas	ñ	'ālf											kaf
በ	bā	ተ	tā	ቸ	čā	ኀ	hā	ኀ	huā	ነ	nā	ኘ	ñā	አ	'ā	ከ	kā	ከ	kuā
ቡ	bu	ቱ	tu	ቸ	ču	ኀ	hu			ኑ	nu	ኘ	ñu	አ	'u	ከ	ku		
ቢ	bi	ቲ	ti	ቸ	či	ኀ	hi	ኀ	hui	ነ	ni	ኘ	ñi	አ	'i	ከ	ki	ከ	kui
ባ	ba	ታ	ta	ቸ	ča	ኀ	ha	ኀ	hua	ና	na	ኘ	ña	አ	'a	ከ	ka	ከ	kua
ቤ	be	ቲ	te	ቸ	če	ኀ	he	ኀ	hue	ኔ	ne	ኘ	ñe	አ	'e	ከ	ke	ከ	kue
ብ	ba/ø	ታ	ta/ø	ቸ	ča/ø	ኀ	ha/ø	ኀ	hua	ኔ	na/ø	ኘ	ña/ø	አ	'a/ø	ከ	ka/ø	ከ	kua
ቦ	bo	ቲ	to	ቸ	čo	ኀ	ho			ኖ	no	ኘ	ño	አ	'o	ከ	ko		

Writing systems (5/6)

- N'ko script
 - Pure RTL writing (with tones)
 - Designed initially for Manding languages in 1948
 - Increasingly popular
 - Recently encoded in Unicode
 - Sample:

	Ⲁ	ⲁ	Ⲃ	ⲃ	Ⲅ	ⲅ	Ⲇ	ⲇ	Ⲉ
	aw	o	uh	eh	e	a	ah		
ⲉ	Ⲋ	ⲋ	Ⲍ	ⲍ	Ⲏ	ⲏ	Ⲑ	ⲑ	Ⲓ
m	l	k	r	d	ch	j	t	p	b
ⲓ	Ⲕ	ⲕ	Ⲗ	ⲗ	Ⲙ	ⲙ	Ⲛ	ⲛ	Ⲝ
ng	y	w	h	n	ny	f	gb	s	rr

Writing systems (6/6)

- Others of historic interest / contemporary use (not L10N priorities)
- **Vai** – oldest indigenous script in West Africa, used for a minority language
- **Kikakui** – Mende script in WA, still some use
- **Bamum** – Cameroon, many mss.
- **Mandombe** – Recent script designed first for Kongo language (Congo)

What is L10n in Africa?

- Largely content, some discussion of keyboards, a little software localisation
- Driven largely by development concerns (IDRC, OKN, OLPC)
- Modest local interest (FOSS); contributions by expatriate Africans
- Some commercial interest:
 - Microsoft “LIP” in some languages
 - Corel WordPerfect spellcheck in S. African languages

Multilingual enabling

- The “last mile” of i18n
 - Includes input methods, fonts, equipping systems (e.g., in project computers)
- With newer Unicode-aware systems some aspects are taken care of (though users often unaware of this)
- Still questions of fonts & keyboards
 - Particular attention to keyboard layouts & some production keyboards
 - Need for more fonts

Content localization

- How much content in African lang's?
 - Some say not much
 - Some say more than we realize
 - Japanese project to census content
- Nature of the content
 - *About* African languages, e.g. instruction
 - Some development-related content
- Level of demand for more not so visible
- Think beyond static web-based content
 - Some forums, blogs in African languages

Software localization

- Proprietary & open-source software
 - Open-source (FOSS) initiatives
 - Microsoft agreements for localization
- Projects
 - Translate.org.za (Sepedi, Zulu, Afrikaans)
 - KiLinux (Swahili)
 - ICT Translations Uganda (Luganda)
 - Kinyarwanda, other nascent efforts
- “Lower the bar” to facilitate more I10n
- Training & networking for FOSS
 - Also including localization

Mobile devices

- Localization of commands & facilitation of text messaging in African languages
 - South Africa (including predictive text)
 - Ethiopia (prototype in Amharic)
 - Kenya (Swahili)
- Inevitable interfaces with other ICTs
- A key area for attention in African localization & ICT for development?

Other support

- Text-to-speech (Swahili)
- Machine translation (Xhosa, Fulfulde - English)
- Discussion of speech synthesis (Yoruba)

Region: North

- North Africa benefits from advances in Arabic computing in & for the rest of the Middle East
 - Egypt is a leader in this
- FOSS active in Egypt & Morocco
- Tifinagh script encoded in Unicode
 - Morocco's decision to use it in schools was a key impetus to complete this

Region: West

- Nigeria
 - Much focus on keyboards
 - Hausa localization of OpenOffice (early)
 - Microsoft agreement: Hausa, Yoruba, Igbo
 - ALT-I & speech recognition
 - TTS for Ibibio
 - OLPC inclusion of Hausa, Yoruba, Igbo
- Others
 - Burkina Faso – OpenOffice early stages for Jula (Bambara), Moore, Fulfulde; keyboards
 - Senegal – Microsoft agreement on Wolof
 - Benin – Fon
- N'ko script encoded in Unicode
- Content localization limited

Region: Central

- Sango – web content; software planning (OpenOffice)
- Cameroon – experiments with online instruction modules for languages
- Congo & Dem. Rep. of Congo
 - Planning and discussions
- Content – some in Lingala, including on some e-mail lists

Region: Southern

- South Africa
 - Translate.org.za translated OpenOffice into all South African official lang's
 - Cellphone localization (impetus from the companies, notably Nokia); predictive text
 - Microsoft agreements for several lang's
- Other countries
 - Malawi (some talk)

Region: East

- Amharic & Tigrinya localizations in Ethiopia & Eritrea
- KiLinux localization of OpenOffice in Swahili as “Jambo” in Tanzania
- ICT Translations Uganda (Luganda)
- Text-to-speech Swahili in Kenya
- Kinyarwanda OpenOffice is advanced
- FOSS in Madagascar

What seems to be working & not?

- Some local efforts for software L10N
 - Initiative is starting point
 - Official support essential on policy & standardization issues
 - “Volunteer” model has limited potential
- Helpful to have greater communication among initiatives
- Some outside projects can have big impact (attn to methods, however)

Keys to success?

- Local initiatives
 - Training (technical, organisational)
 - Facilitate mutual aid
 - Tools (e.g., Pootle)
 - Information resources
 - Funding for localisation (from where?)
- Broader vision & planning
 - Local initiatives “can’t see the forest...”
 - Cross-border languages
 - Strategy for outreach to areas w/ potential
 - Need to attend to the localisation ecologies
 - Widen network to other “stakeholders”

Challenges

- Localisation not a policy yet for most of Africa – can this be changed?
- Negative attitudes about African languages among many donors and even some Africans – how to deal with?
- Basic ICT and literacy issues – can localisation work be linked with overcoming these?

Thank you

- Questions...
-