

Information Poverty due to Linguistic Diversity in South Asia and Sub-Saharan Africa

**Policy-recommendations for making online
content available in local languages**

Ranjit Goswami

IIFT, Kolkata

16th December, 2007,
Chennai, India

Content and Convergence

Agenda

- Importance of online content in local languages
- The revolution in online content /information
- Linguistic Diversity (LD) in SA and SSA
- LD and Per Capita GNI in SSA
- The inequality in online content in SA & SSA
- Policy Alternatives
- Policy Recommendations

SA = South Asia
SSA= Sub-Saharan Africa

Information, education, content breaks the vicious cycle

- Relationship well proven
- Role of media, mass media
- Limitations of mass-media in resource deficit, linguistically diverse SA and SSA (non-web-based)
- Opportunities that online content presents
- Informed society thro' online content
 - Low cost of storage (Moore's law), software-based translation
 - Anywhere, anytime & easy usability; ↑ accessibility
- Barriers essentially literacy, language, relevance (Wendy & Francisco, 2000; Rao, 2000)

Information, content everywhere, but not for ones who need it most...

- The online content revolution... 'How Much Information' (Lyman, P, 2000...Berkeley project)
- 'Using What You (We) Know' collectively...(BW, 2006) – example of Galileo
- Monopolization of English as online content, particularly for South Asia (SA) and Sub-Saharan Africa (SSA).
- The 3rd line of division...(remember AJ speech in CPRsouth1...knowing English as a passport to a job in India...true for SA & SSA)

The paradigm shift in information cost...online content is free

- With connectivity, one gets access to almost all the information one needs
 - The barriers for ‘knows’ and ‘know nots’ diminish with connectivity, and language skills.
- The positive side: connectivity is on the rise...
- The negative: What about language skills for SA and SSA?
- Physical connectivity (the focus so far) vis-à-vis mental connectivity (neglected area)...
 - online relevant content in local languages
- And that’s our research problem...

Redefining universal access

- Physical reach – availability (devices), accessibility (price points), skill-sets (easiest for search engines)...all +ve – ‘Hard Infrastructure’
- The language divide – another ‘last mile’ – ‘Soft Infrastructure’
 - So long ignored, as focus more on physical connectivity.
- SA and SSA linguistically fragmented with **poor literacy, poor primary education, high poverty**, less resources.
- Online content can act as a social tool – language barrier?

LD – a hurdle in underdeveloped nations?

- 70% of people (1.2 billion) living below \$1-a-day in SA & SSA (Pigato, 2001)
- Less than 60% adult literacy (UNESCO, 2006)
- 35-40% students drop out in primary level itself
- Less than 8% of Indians complete ten years in formal schooling, <4% for 12th level (Wikipedia)
- Similar situations in other nations of SA & SSA
- Common cause – Linguistic Diversity?

LD and per-capita GNI in SSA

- Studied 10-most linguistically diverse nations from SSA
- These nations had an average per-capita GNI (PPP) of \$1305 against \$2031 for SSA (excl. South Africa).
- Eight of the nations had less figures
- Trend naturally follows for Internet usage as well (1.83% against SSA score of 3.6%, incl. South Africa)

Linguistic Diversity & Content Inequality

- 10% people in SA and SSA read English (for SSA, other European languages), 0.2% as 1st language#
- 22 major languages in India; 4 provisional, two regional, 2 official & 1 national language in Pakistan; 2 major languages in Sri Lanka
- Swahili, major SSA language has only 5-10 million native speakers, 80 million speakers (hardly 10% of SSA population)
- Cameroon, the most linguistically diverse nation of SSA, had only 10.4% of its population speak in three of the largest native language groups.
- 25 web-pages/Internet user: 14-20 are in English#

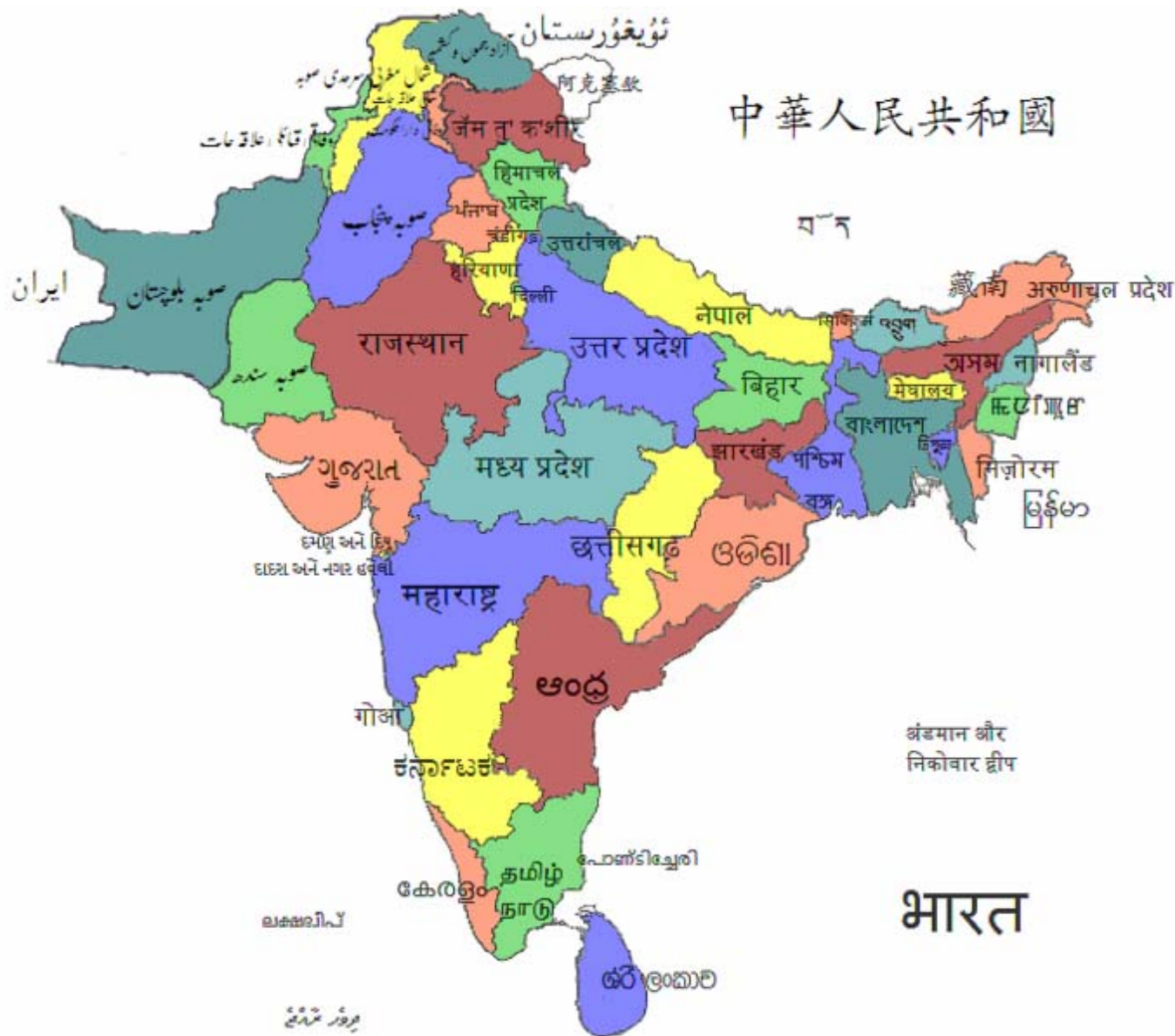
#These numbers varied as difficult to research these dynamic areas. We present a consensus approach.

16th December, 2007,

Chennai, India

Content and Convergence

Linguistic Diversity in SA



16th Dec
Chennai, India

Source: Wikipedia

Simply put...

- No native language from SA or SSA in top 18 Internet languages (netz-tipp-de studies, 2002)
- Two South Asian Languages in top ten languages of the world (including secondary speakers, one in top 5).
- In contrast of – around 1% of Korean speaking people contribute 3% of Internet users, 1.5% of online content
- Chinese growth in Internet (400%) driven by local language content (English language 157%)
- 56% online content in English, followed by 7.7% German, 5.6% French, and 4.9% Japanese (netz-tipp-de studies, 2002, we believe Chinese language content to be significant now and going forward)
- Roughly 155 web-pages/native language English speaker, 104 (25) web-pages/English language speaker (including 2nd language) whereas much less than 1 web-page/Bengali or Hindi speaker (#figures varied in different sources, our ballpark calculations)
- Compare with maximum income disparity of 1:50 order for these two regions against developed world standards
- Where are the content for nearly 40% people of the world?

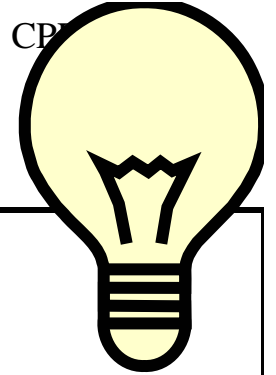
User Generated Content does little better...

- Wikipedia contains 8.2 million articles in 253 languages, % of English much less at 24%
- Still only 6 languages from SA in top 70 (more than 10K articles)
- Only Swahili from SSA ranked 88th with 5863 articles (0.07%)
- Lingua project (globalvoicesonline.org) inadequate
- The hope in Web 2.0, UGC, video content, translation software, blog-support in local languages, local-language support in search engines, etc. (e.g. Baidu in Chinese 61% market-share, Google less than 1/2 of it)

The problem...the solution

- 80%-90% of health and food-related institutions didn't support multiple languages
 - For pandemic diseases prevalent in SA & SSA
 - Preference to search in local language even amongst well-off sections (Singh et al, 2007)
- Limitations of earlier media was a handicap
- Advantages of online content –limited by physical reach and language barrier – better option?
 - Physical connectivity has been improving over the years (highest growth rate in SA & SSA, 2000 onwards)
 - Language barrier remains and will be the only hurdle

The Policy Choice



Educate people in English/other foreign languages (for SSA)	Educate people in mother-tongue. (basic/primary level)
Use existing content...no further action needed	Partner, convert and archive content in local languages with SW support.
Present level of literacy, quality of primary education	Generate content where absolutely necessary
Drop-out rates...looks possible?	Use online content as a social tool...let there be light

Don't reinvent the wheel!

- Shortages of resources in SA and SSA
- Focus on content availability in local languages...
 - Too much content...too much information already exists and are added daily, in other languages
- No need to generate content, unless absolutely necessary.
- A combination of initiations, partnerships & innovations to aid market forces
- Commercially viable option (ad-based revenues) as input costs insignificant, potential huge market (e.g. of China, Korea, Japan, Russia, Middle-East)

Thank You

ranjit.goswami@gmail.com