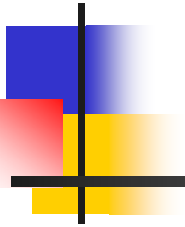


PAN Localization Phase I: Successes, Challenges and the Future



Mumit Khan
Center for Research on Bangla
Language Processing (CRBLP)
BRAC University



Highlights

- Research center dedicated to Bangla language processing
- Technology development objectives met; in general exceeded original goals
- Human resource development on target
 - Gender balance not satisfactory
- Groundwork established for local and regional collaboration
- Challenges in affecting national policy



Institutional development

- CRBLP - the only research center in Bangladesh dedicated to language processing
- Funding from PanL10n, Microsoft Corporation and BRAC University
 - Pursuing concurrent funding from EU through the Asia Link program
- Alumnae going on for higher studies in NLP/Computational Linguistics
- Significant research output



Technology Development

- Original deliverables
 - Spelling checker
 - Optical character recognizer (OCR)
- Extended deliverables
 - Bangla rich text editor for web authoring
 - Tagged lexicon
 - News corpus and analysis
 - POS tagger, n-gram model, ...
 - Encoding converters, text processing utilities



Human resource development

- 11 full-time staff trained in Phase I
 - 1 pursuing PhD
 - 3 on the way to MS/PhD
- Significant turnover in the first year
- Recent collaboration with Dhaka University Linguistics department
- 3 trained during Summer School



Policy development

- Submitted proposals to Bangla Academy for
 - Collation standard (UCA)
 - Electronic version of its standard dictionary
- Significant presence in Bangladesh Computer Council
- Close collaboration with Prime Minister's Office



Challenges

- Human resource
 - capacity development
 - retention (DU collaboration will add MS/Mphil opportunities)
 - Lack of gender parity
- Affecting national policy. Perhaps no longer?



“Side Effects”

- 90% of research papers published from CRBLP
- First ever (and the only) course on NLP taught in Bangladesh
- Excitement and interest in a new generation of students to work on NLP
- Microsoft Vista and Office localization
- Open source development catching on?



Future

- Open source advanced applications that will directly impact end users
- Training end users via partner organizations
- Capacity co-development through collaborations