Building Capacities in Human Language Technology for African Languages

'Tunde ADEGBOLA

African Languages Technology Initiative (Alt-i)
Ibadan, Nigeria.

Supported by:

Tiwa Systems Ltd.,
Bait-al-Hikma,
Open Society Initiatives for West Africa,
International Research Centre (IDRC).
Aim of this presentation

➢ Describe efforts on African language technology
  ➢ Focus on work at African Language Technology Initiative (5-years: 2003 to 2008)
➢ State challenges and opportunities for African language technology
➢ Present proposal for accelerating the development of African language technology
State of African Language Technology

- Relatively recent; expanding
- Efforts in South Africa
  - motivated and guided national policy
  - private sector and public organisations
  - semi-government institutions
- Efforts in other parts of Africa are based on private initiatives
- Encouraging International assistance
  - Mainly from Europe
South African Effort

➢ Based mainly in 7 universities:
  ➢ University of Cape Town
  ➢ University of Limpopo
  ➢ University of the North West (Potchefstroom)
  ➢ University of Pretoria
  ➢ University of South Africa
  ➢ University of Stellenbosch
  ➢ University of the Witwaterstrand (Johannesburg)

➢ Semi-Government institute
  ➢ Meraka Institute
  ➢ Human Language Technology Unit (Under department of Art and Culture)
Other efforts in Africa

➢ West Africa
   ➢ Only one private organisation: The African Language Initiative (Alt-i)
   ➢ Individual (O.A. Odejobi)

➢ East Africa
   ➢ The Djibouti Centre for Speech Research
   ➢ Technobyte Speech Technologies (Kenya)
   ➢ Individual (Wanjiku Ag'ang'a, Peter Wagacha)
Efforts in other parts of the world

- AflaT
- Outside Echo Project (UK):
  - Local language speech technology Initiative
- West African Language Documentation Project (Germany):
  - University of Bielefeld and University of Uyo (Nigeria)
- Other small activities:
  - E.g. In USA, Yoruba-English machine translation at St Mary's College of Maryland
History

- Started in 1975 but became more focused in 1985
- By Electrical engineers and physicists
- Realises the importance of linguist in 2001 and incorporate linguistic experts
- Based at Ibadan, Nigeria
- Efforts primarily focused on Yoruba
- Initial connection with the academia was hampered by bad economy
  - This has improved, but interdisciplinary efforts still low
Activities

➢ Includes research and development in the following areas:

➢ Automatic speech recognition
➢ Text to speech synthesis
➢ Machine translation
➢ Yoruba spelling checker
➢ Automatic diacritic application
➢ Localisation of Microsoft Vista and Office
➢ Assistance to Universities
➢ Education
Started in 2001

Approaches ASR through the use of tone information (similar to talking drum)

Findings

Tone-guided search of the recognition space produce improved accuracy and speed

Results include:

- A PhD Thesis
- Yoruba speech recognition resources
- Efforts continuing (funded by OSIWA)
Text-to-speech (TTS) Synthesis

➢ Started in 2002

➢ Results

➢ Our associated (OA Odejobi) researched into prosody modelling for Yoruba TTS

➢ Used an innovative modular holistic approach which integrates: Relational tree and fuzzy logic

➢ Book on the technique and how it can be extended for other African languages published (available at Amazon)

➢ Funding yet to be obtained for sustaining this work
Focus on translation of language spoken in Nigeria to English
- Igbo-English
- Yoruba-English
Efforts of student volunteers from Department of Linguistics and African Languages and Africa Regional Centre for Information Science
Funding yet to be obtained for sustaining this effort
Yoruba spelling checker

- Work as part of African Network of Localization
- Developing spelling checker for Open Office
  - Based on Hunspell software (Nemeth Laszlo)
    - Hunspell cannot accommodate all Yoruba morphology rules; separate codes were developed to handle this.
- Computational study of Yoruba morphology
- Involves staff and Students of Department of Linguistics and African Languages at the University of Ibadan

Results

- ~ 5000 Yoruba root words
- 100 highly productive affix rule
- Working (but limited) spelling checker
- Funded by International Research Center, Canada
Automatic diacritic application

➢ Aim is to generate automatic text tone maker for accurate Yoruba orthography
➢ By product of Yoruba spelling checker project
➢ Uses the Bayesian learning approach
➢ Uses corpus produced in the IDRC
➢ Funding yet to obtained for this project
Localization of Microsoft

- Microsoft appointed Alt-i as moderator for localising its Vista and Office Suite
- Working on Hausa, Igbo and Yoruba
- Project progressing
Assistance to Universities

➢ Teaching of PG students at University of Ibadan
➢ Supervision of postgraduate projects at African Regional Centre for Information
➢ Provide facilities for many PhD and research students
➢ Provide facilities and support staff and students from a number of universities in Western Nigeria
➢ Collaborate with a number of organisations (e.g. WALS, LAN & YSAN)
Education and outreach

➢ Seminar
  ➢ In 8 Nigerian universities

➢ Workshop and conferences
  ➢ For scholars in Linguistics, physics, computer science, etc.

➢ Cross-disciplinary studies
  ➢ Encourages and support knowledge and skill sharing
Observations

- Intellectual resources are available in the universities.
- Lack of awareness hampers focussed and organised effort and hence progress.
- Sentimental attachments to departmental traditions prevent positive engagement.
- Importance and role of linguistics in language technology development not given adequate recognition.
- Inappropriate admission criteria and limited curricular...
Recommendations

➢ Intensive and sustained awareness building programmes on language technology
➢ Review of admission criteria and curricular to encourage and sustain students interest
➢ Employ modern technique for management of learning resources
Advocacy
- Identify and develop policy thrust - encourage development of African language
  - Accelerate the development of African language technology
  - Produce lecturer, researchers and other experts
  - Raising awareness in secondary and tertiary institution

Service
- Develop man power through graduate training
- Support from international scholars will be sought
- Develop product that will draw attention to language technology
Conclusion

➢ Development of African language technology is in embryonic state
➢ Apart from South African Efforts, no coherent efforts in Africa
➢ National language policies do not address language technology appropriately
➢ Low level of the awareness of the benefits of language technologies
➢ Interdisciplinary and multidisciplinary efforts are required
Thank you

Suggestions and Question?