Session 3 : Scripting Operational Experiences
Highlights & Conclusions

First Speaker: Michael Everson (Evertype)
"Scripts Encoding Initiatives – Africa"

Second Speaker: Ivan Guzman (IGRAL)
"Experience with Language Implementation in ATAMIR"

Third Speaker: Nimaan (Universite d'Avignon)
"Speech Mining to make African Oral Patrimony Accessible"

Chair/Rapporteur Tarcisio Della Senta

UNDL Foundation

o   UNICODE covers the economic feasible scripts, however, encoding scripts remains an important issue for many languages for strong cultural and political reasons.

o   Examples shown of N'Ko, Solomane Kante, not only sets, also keyboard design. N'ko keyboard design is (an example to be generalized. Stage II). N'ko font design, Unrectified glyph – rectified glyph design.

o   Why are so many scripts missing?
   •   Scripts missing: not many experts in the world, limited resources
   •   Recommendation: find sponsors for scripts **!!**

o   What it means when a script is encoded? www.ethnomed.org (example). Benefits of encoding scripts: connected to the world; there are also economic and cultural negative consequences of lack of encoding.

o   Important to make universal

o   Everson does not agree to stop the process of new encoding. Work in Unicode should not stop with a fixed, arbitrary deadline. There is a pressure to finish the effort in UC/ISO.
   •   He made 200 proposals in the last decade to Technical Committees in UNICODE.

o   Role of SEI, available for free to the target communities thanks to UNESCO's funding; participate in standards committees, advocacy, promotion.
   •   Other groups involved : SIL, Language schools, minority language groups
   •   SEI: 12 scripts, 8 scripts in progress.

Geneva, 9-11 May 2006

2

o ATAMIRI, a native language from Bolivia used as an intermediary for translation.

o Allows for automatic translation across languages. Some 8 languages supported, more being added. Examples are shown.

o Lexicographic project is done with volunteers. Promotes lexicographical discussion forums. People rates the quality of translations.

o asks people for rating on the quality of the translation

o Tasks ahead: development of dictionary (for terminology), improve codification in lexicographic databases, programs to solve well known ambiguities

o 72 translation directions currently implemented; major part (90%) of anomalies can be solved in the framework of the current ATAMIRI design, except for some very special cases.

o Has a prototype multilingual IM software

o Various proposals at end of paper. Create an ATAMIRI Language Engineering network (ALEN) for the continuous evolution of the translation engine

o Funding commitment is needed to enable the work to be done

o Keep it free and open source

o Most African countries follow an oral tradition system to transmit their cultural, scientific historic heritage through the centuries. Development of speech processing technologies

o Ancestral knowledge accumulated during centuries is today threatened of disappearing. African countries have a strong oral tradition

- African countries should become aware of the need to save its heritage
- It should also be made more accessible

o The presentation describes the first steps for automatic transcription and indexing of African oral tradition heritage.

o The focus is on Djibouti's cultural heritage. Djibouti Language is tonal, with 22 consonants and 20 vowels. Consonants have a phonetic structure. An Acoustic Speech Recog nition technology was used for the Somali language, and applied to a an audio corpus, and a textual corpus.

o On the transcribed oral corpus, indexed database is created and then speech mining technique is used to retrieve data from the oral archives.

o Transcription tools are needed

- Automatic system for conversion between speech and written words
- Objective is not to make 100% automatic transcription, but to save the information.

Session 3: General Conclusions

o Speakers highlighted elementary problems that persist in many local languages, such as scripts encoding, transliteration of traditionally oral languages, and lock them far in the other side of digital divide.

o The solution of these problems lies in the hands of some dedicated scholars, NGOs, and international organizations concerned with cultural heritage, in particular UNESCO.

o Funding and political support is the most common request.

o They call for support and openness from the "big managers" of the Internet protocols and technologies to put in place the "basics elements" for accessing it, such scripts encoding.

o A general call for collaboration among different stakeholders particularly of those working of less "digitalized" languages.

o Technologies are needed for automatic system for conversion between speech and written words

o Create an ATAMIRI Language Engineering Network (ALEN) dedicated to further develop and exploit ATAMIRI's technology in its full potential.

o Transfer ATAMIRI technology to ALEN with the system creator as on of its stakeholders. The stakeholders have the option to adopt an open source model for the software.

o Assistance from NGO, UN organizations and from the private sector in order to be able to accomplish encoding endangered scripts tasks.