UNITED NATIONS EDUCATIONAL,
SCIENTIFIC AND CULTURAL ORGANIZATION

**The UNESCO's contribution for the inclusion of languages and local content in cyberspace: from oral languages to digital objects**
**Claudio Menezes**
**Senior Programme Specialist**

**On the occasion of the**

**ITU-UNESCO**
**Global Symposium on the Multilingual Internet**

**International Conference Centre of Geneva**

**Switzerland**

**9 May 2006**

From oral language to digital objects: the UNESCO contribution to a multilingual cyberspace

Cláudio Menezes
Senior Programme Specialist
Information Society Division
Communication and Information Sector
UNESCO – Paris
email: c.menezes@unesco.org
http://www.unesco.org/webworld

## 1.      Introduction

In a world that is increasingly "digital", there is an unprecedented opportunity in the form of those ICTs that are the medium for transmitting knowledge and education. The participating process in social and political life also encompasses a "digital" component of which language is part of. These are some of the reasons why UNESCO addresses "multilingualism in cyberspace" and particularly on the InterNet from different angles: as a gateway for quality education, to freedom of expression and as an element of an strategy to preserve and promote language diversity.

The linguistic diversity has been present in different regions of the globe over the centuries. However, this linguistic diversity is still far from being reality in the digital world.

As per a study developed by "Ethnologue" (Table I), the linguistic diversity index scores by region show that the United States present the lowest level (0,7809) of linguistic diversity, Africa being the continent with higher level of the same score (185,6836). There are indications that global linguistic diversity is currently in decline and it has been so for a long time.

**Table 1**
**Linguistic diversity index scores by region**

| Region | Languages | Diversity index | Prop. Of world total |
|---|---|---|---|
| USA | 170 | 0.7809 | 0.0020 |
| N Am. (incl. USA) | 248 | 3.3843 | 0.0086 |
| E Asia | 200 | 4.4514 | 0.0112 |
| W Asia | 159 | 26.1539 | 0.0659 |
| SC Asia | 661 | 29.8093 | 0.0752 |
| S America | 930 | 30.5007 | 0.0769 |
| Europe | 364 | 32.4369 | 0.0818 |
| SE Asia | 1317 | 37.6615 | 0.0949 |
| Oceania | 1322 | 46.5653 | 0.1174 |

| Africa | 2390 | 185.6836 | 0.4681 |
|---|---|---|---|

Another element for alarm: according to some estimates, nearly half of the world's languages will be extinct by the year 2050 (see "Measuring Linguistic Diversity on the Internet", by John Paolillo and alii, 2005, UNESCO Institute of Statistics).

Information and communication technologies play a double role in the recent linguistic world transformations: on one side, they can be an important vehicle for communication among different linguistic communities; on the other hand, ICTs can also be factor for strengthening the marginalization of languages in the cyberspace. There are more than 6,000 languages, out of which 12 languages respond to 98% of the webpages on the Internet. English with 72% of webpages, is the dominant idiom accordingly to an experiment conducted by O'Neill, Lavoie and Bennet for the 'On Line Computer Library Center – OCLC' in 2003.

There is therefore a need to improve this situation by finding a better balance on linguistic diversity in the digital world through: a) inclusion of new languages; and b) inclusion of new content in under-represented languages in the cyberspace.

The inclusion of languages and local content on the digital world requires the development of a technological environment to allow the creation and the access to "digital objects" (text, sound and images on digital form). The creation of this "enabling environment" involves some preparatory work to make sure that an oral language can be mapped into digital objects. In general, this process includes the following general components:

| | |
|---|---|
| a) | Identification of a written form of an oral language (or dialect) |
| b) | Identification of scripts and semantic rules for a particular language (for example, many languages use the Cyrillic alphabet) |
| c) | Mapping the symbols into a digital table (this work is prepared by experts and submitted to UNICODE Consortium); |
| d) | Establishment of standardized fonts of characters. |
| e) | Hardware manufacturing and software development. |

Once established these preliminary and preparatory works, hardware and software developers are able to manufacture equipment and software systems that allow the use of a language in the digital world. While the development of an enabling environment is a feasible task, it can involve investments that make difficult the inclusion of many languages in the digital world (the development of linguistic versions of operating systems and user interfaces in as many languages as possible is particularly important).

The "enabling environment" is a pre-requisite for the promotion of "local digital content" in the cyberspace.

There are therefore two major challenges before the national and international communities:

a) to ensure appropriate investments both from the private and the public sector enabling the inclusion of as many languages as possible of new languages in the cyberspace
b) to ensure that a multilingual and culturally diverse cyberspace will arise, mapping the actual real world 's linguistic diversity into the cyberspace.

All in all, the challenge before the international community is to overcome the gigantic obstacles to make sure that a multilingual and culturally diverse cyberspace will be built up.

## 2. **The UNESCO contribution to multilingualism in the digital world**

Several normative texts and declarations have testified to a growing awareness of the importance of language issues at the international level (e. g. the "Declaration of Vienna" of the World Conference on Human Rights (1993), the Resolution 56/262 of the General Assembly of the United Nations and the report of the Secretary-General of the United Nations at its fifty-eighth session (2003) and the "World Summit on the Information Society" in both Geneva and Tunis phases)
.
Under request of its Member States, during the two phases of the World Summit on the Information Society, UNESCO has been consistently advocating the need for building up inclusive and pluralistic knowledge societies, based on four principles: freedom of expression, education of quality for all, cultural and linguistic diversity and universal access to information and knowledge, specially of information in the public domain.

The UNESCO Member States representatives also approved the "Universal Declaration on Cultural Diversity", in 2001. They also approved the "Convention for the Safeguarding of the Intangible Cultural Heritage". Most recently, in October 2005, the 33rd General Conference approved a "Convention on the protection and promotion of the Diversity of cultural expressions". The "Universal Declaration on Cultural Diversity" includes a number of lines of actions, three of which address directly questions of digital media and information technology:

"*9. Encouraging "digital literacy" and ensuring greater mastery of the new information and communication technologies, which should be seen both as educational discipline and as pedagogical tools capable of enhancing the effectiveness of educational services.*
*10. Promoting linguistic diversity in cyberspace and encouraging universal access through the global network to all information in the public domain.*
*11. Countering the digital divide, in close cooperation in relevant United Nations system organizations, by fostering access by the developing countries to the new technologies, by helping them to master information technologies and by facilitating the digital dissemination of endogenous cultural products and access by those countries to the educational, cultural and scientific digital resources worldwide (UNESCO, 2001, pp. 60)*"

These principles and lines of action establish values for evaluating attributes of the information society in ethical terms and goals for its development. However, they do not provide sufficient insight into the possible causes for any biases that might exist. On this topic, an interesting collection of papers on "Measuring Language Diversity on the Internet"[1],[2] published by UNESCO Institute of Statistics addresses the different biases on monolingualism on the Internet.

Multilingualism in cyberspace has been present in UNESCO's programme since 2002, when the Committee of the Intergovernmental "Information for All" Programme decided to submit to the General Conference a "Draft Recommendation on the Promotion and Use of Multilingualism and Universal Access to Cyberspace".

Since then, three major elements have constituted UNESCO's approach on multilingualism in the digital world:

| |
|---|
| 1)   the approval of recommendations and resolutions on policies by UNESCO's Member States (through the General Conference and Executive Board); |
| 2)   the elaboration of technical documents, guidelines and studies supporting policies and awareness initiatives; |
| 3)   pilot projects and the dissemination of on-line resources and multilingual tools. |

There are a detailed number of publications, brochures, studies and technical documents on each of the above three topics, available on paper or in digital form[3]. Of utmost relevance to the policies approved by UNESCO Member States, I wish to mention: a) the "Recommendation on the Promotion of Multilingualism and Universal Access to Cyberspace", approved by the 32nd General Conference, in 2003; and the "Report by the Director General on an Overall Strategy for an Intersectoral Programme for Languages at UNESCO" (170 EX/20, approved by the Executive Board). A more complete list of pilot projects, technical documents, guidelines and studies is available on the "Initiative B@bel" brochure[4], as well as from the multilingualism section of UNESCO website (http://www.unesco.org/webworld/multilingualism). Examples of these UNESCO initiatives are: Education in a Multilingual World (policy paper); National Terminology Planning Guidelines (technical report); Thematic Edition of the International Journal on Multicultural Societies "Language on the Internet"; Sample Script Implementations; Guidelines for Development of Writing Systems Support; Complex Data Exchange and Format; Intellectual Property Writing

---

[1] English Version: http://unesdoc.unesco.org/images/0014/001421/142186e.pdf
[2] French Version: http://unesdoc.unesco.org/images/0014/001421/142186f.pdf
[3] UNESCO WSIS Virtual Stand, http://portal.unesco.org/ci/wsis/tunis/stand/index.php?Action=showHome
[4] http://portal.unesco.org/ci/wsis/tunis/stand/index.php?Action=showThemes&id=298&theme=239

Systems; Handbook for Linguistic Data-Collection by Non-Linguists. Further examples are: Report "Measuring Linguistic Diversity on the Internet"; The Multilingual Content Management Platform (MLCM), SILA – A Multilingual Web Browser, Modified version of WorldPad, Graphite Edit Control, Multilingual Cultural Heritage Knowledgebase, Internet Sound Record Library of Endangered Caucasian Languages, Abkhazian-Georgian Dictionary Project, Ethiopic Computer Standards and UNESCO Atlas of Languages in Danger of Disappearing.

The most recent UNESCO initiatives on multilingualism in the digital world are related to the World Summit on the Information Society (Tunis, 16-18 November 2005). A preparatory WSIS thematic meeting on "Multilingualism for Cultural Diversity and the Participation of All in Cyberspace" was organized by UNESCO in Bamako, Mali, on 6-7 May 2005 (see "Report and Conclusions" at http://unesdoc.unesco.org/images/0014/001402/140208E.pdf). In addition, UNESCO published during the Summit the UNESCO World Report "Towards Knowledge Societies" (available from http://unesdoc.unesco.org/images/0014/001418/141843e.pdf), which also addressed the challenges of multilingualism in the digital world. The topic on multilingualism in the digital world is part of the WSIS Plan of Action (Action Line C8 – Cultural diversity and identity, linguistic diversity and local content) and is reflected on the "Tunis Commitment" and on the "Tunis Agenda for the Information Society"[5].

## 3.    Forthcoming actions and Conclusions

In UNESCO, the forthcoming actions to promote language diversity in cyberspace will continue in the framework of the approved programme for 2006-7, including with particular emphasis to the African Year of Languages, as declared by the African Union. Initiatives such as the World Directory of Institutions and Experts, Network of Institutions Specialized in Multilingualism in Cyberspace, Report on the national implementation of the "Recommendation on the Promotion and Use of Multilingualism and Universal Access in the Cyberspace" and the proposal to United Nation General Assembly to declare 2008 as the 'International Year of Languages' are planned for this biennium. Two new initiatives have been supported by UNESCO, namely: a) the World Network for Linguistic Diversity (redili@funredes.org), launched on 21 February 2006 (International Mother Language Day) by the African Academy of Languages and partners in UNESCO headquarters; b) the present ITU-UNESCO "Global Symposium on Promoting the Multilingual Internet" (Geneva, 9-11 May 2006, http://www.itu.int/ITU-T/worksem/multilingual/).

Can these UNESCO initiatives contribute to make "multilingualism in the digital world a reality"? Can they help in the construction of a "multilingual Internet"? I hope so. Can it make reality the utopia of multilingualism in the digital world? It is difficult to reply.

---

[5] http://www.itu.int/wsis/documents/doc_multi.asp?lang=en&id=2266|2267

In any case, the challenge of making more than 6,000 languages present in cyberspace is a multi-stakeholders one. International organizations like UNESCO have a role to play. National governments, the private sector, international banks, NGOs, academy are very important stakeholders to make it happen. Given the magnitude of such a challenge, a collective multi-stakeholder process should be put in place. The production of local digital content and digital objects requires the carrying out of an 'invisible work' to make sure that the enabling environment will facilitate cultural and linguistic diversity in the digital world. It will then be possible to complete the whole cycle from oral languages to digital content.