LINGUISTIC AND CULTURAL DIVERSITY IN CYBERSPACE

Proceedings of the 3rd International Conference

Ministry of Culture of the Russian Federation
Federal Agency for Press and Mass Communications
of the Russian Federation
Government of the Republic of Sakha (Yakutia)
Commission of the Russian Federation for UNESCO
Russian Committee of the UNESCO Information for All Programme
Ammosov North-Eastern Federal University
Interregional Library Cooperation Centre

Linguistic and Cultural Diversity in Cyberspace

Proceedings of the 3nd International Conference
(Yakutsk, Russian Federation, 30 June – 3 July 2014)

Moscow
2015

Linguistic and Cultural Diversity in Cyberspace. Proceedings of the 3rd International Conference (Yakutsk, Russian Federation, 30 June – 3 July, 2014). – Moscow: Interregional Library Cooperation Centre, 2015. – 408 p.

The book includes communications by the participants of the 3rd International Conference on Linguistic and Cultural Diversity in Cyberspace (Yakutsk, Russian Federation, 30 June – 3 July, 2014), where various aspects of topical political, philosophical and technological challenges of preserving multilingualism in the world and developing it in cyberspace were discussed. The authors share national vision and experience of supporting and promoting linguistic and cultural diversity, express their views on the role of education and ICTs in these processes.

The authors are responsible for the choice and presentation of facts and for the opinions expressed, which are not necessarily those of the compilers.

Contents

PREFACE

The Third International Conference on Linguistic and Cultural Diversity in Cyberspace took place in Yakutsk (Russian Federation) on 30 June – 3 July, 2014.

It is a significant contribution made by the Russian Federation in the activities of UNESCO, which considers the preservation of linguistic and cultural diversity as one of its top priorities.

The conference is also Russia's new contribution in the implementation of the UNESCO Intergovernmental Information for All Programme (IFAP) – one of UNESCO's two major programmes in the field of communication and information.

The event was organized by the Russian Committee of the UNESCO Information for All Programme, the North-Eastern Federal University in Yakutsk, and the Interregional Library Cooperation Centre in cooperation with the Commission of the Russian Federation for UNESCO. Financial support was provided by the North-Eastern Federal University, Government of the Republic of Sakha (Yakutia), Ministry of Culture of the Russian Federation, Russia's Federal Agency for Press and Mass Communications, and UNESCO.

The conference gathered representatives from almost 50 countries of diverse regions of the world – leading experts, workers of culture, scientists, educators, politicians and diplomatic officials of Albania, Argentina, Austria, Azerbaijan, Belarus, Botswana, Brazil, Bulgaria, Central African Republic, China, Colombia, Czech Republic, Dominican Republic, Ecuador, Estonia, Finland, France, Hungary, India, Israel, Italy, Japan, Kazakhstan, Kyrgyzstan, Latvia, Macedonia, Moldova, Netherlands, Nigeria, Oman, Peru, Poland, Republic of Korea, Republic of Maldives, Russian Federation, Rwanda, Slovakia, Sri Lanka, Sudan, Sweden, Syria, Thailand, Togo, Turkey, UK, USA. Over half of the participants were nominated by national governments.

Conference participants were glad to note that it attracts more and more attention worldwide. De facto, it has become the world's major forum for discussing topical problems of languages preservation and their development in cyberspace.

The First Yakutsk International Conference on Linguistic and Cultural Diversity in Cyberspace in 2008 gathered representatives of 15 countries – and for the Russian Government and UNESCO it was a big success. It brought up the theme in Russia and became the first event on this topic within IFAP and UNESCO.

The Second Conference in 2011 welcomed participants from 33 countries.

Both conferences attracted great international attention and led to the adoption of important international instruments – the Lena Resolution "On Cultural and Linguistic Diversity in Cyberspace" and the Yakutsk Call for Action "A Roadmap towards the World Summit on Multilingualism".

The Lena Resolution, the final document of the first conference, has received international recognition as the first document structuring the problematic situation in the field of multilingualism promotion and identified all stakeholders. It is currently being cited in research and formal documents of international organizations. The second conference conclusions and final document were discussed at the UNESCO General Conference in 2011. Both conferences' proceedings are published in printed and digital form in Russian and English.

Among the important outcomes of the first two conferences on Linguistic and Cultural Diversity in Cyberspace are the expansion of professional contacts and the establishment of friendly relations between leading experts from different countries. Fruitful partnerships have been established between the UNESCO Intergovernmental Information for All Programme, its Russian Committee and the MAAYA World Network for Linguistic Diversity, headed by Adama Samassekou, Chair of the Preparatory Committee of the World Summit on the Information Society. In 2010 the Centre to Advance Multilingualism in Cyberspace was opened under the North-Eastern Federal University with the support by the Russian IFAP Committee and the UNESCO Moscow Office. Awareness of the importance of issues of multilingualism preservation and development in cyberspace was raised at different levels, primarily within UNESCO itself. On Russia's initiative multilingualism in cyberspace was proclaimed the sixth priority of the UNESCO IFAP and a special IFAP Working Group was created.

All this has led to an even greater interest shown to the Third conference all over the world.

The Conference Opening Gala took place in the Government House of the Republic of Sakha (Yakutia) and its four working days included two plenaries and eight sessions of four sections:

- ICT for linguistic and cultural diversity in cyberspace;
- Socio-cultural aspects of linguistic diversity in cyberspace;
- Preservation of linguistic and cultural diversity in cyberspace: national vision and experience;

- Education for preservation of linguistic and cultural diversity in cyberspace.

Sixty five papers were delivered by the participants.

Russian version of the analytical digest *Net.lang: Towards multilingual cyberspace* was presented at the conference. The book was initially published by the MAAYA World Network for Linguistic Diversity in English and French with UNESCO's support. Most authors of the book took part in the Third conference and had also contributed to the two previous ones.

All conference participants received an impressive set of materials in Russian and English on the issues of linguistic and cultural diversity in cyberspace published by the Russian IFAP Committee and the Interregional Library Cooperation Centre.

These publications formed the basis of a book exhibition opened at the Conference.

Conference cultural programme entailed participating in the Yakut national celebration Ysyakh, visiting the Lena Pillars Nature Park, inscribed on the UNESCO World Heritage List, the Permafrost Kingdom Museum and the Mammoth Museum, and also the Arctic Innovation Center of the NEFU.

In conclusion of their work, participants of the Conference adopted its final document – Yakutsk Declaration on Linguistic and Cultural Diversity in Cyberspace.

Evgeny KUZMIN
Co-Chair, Conference Organizing Committee
Vice-Chair, Intergovernmental Council,
UNESCO Information for All Programme (IFAP)
Chair, Russian National IFAP Committee
Chair, IFAP Working Group for Multilingualism in Cyberspace
President, Interregional Library Cooperation Centre
Member, Commission of the Russian Federation for UNESCO

GREETINGS TO CONFERENCE PARTICIPANTS

Greeting by Getachew Engida,
UNESCO Deputy Director-General

Ladies and Gentlemen,

Dear Friends,

I am pleased to welcome you to this *3rd International Conference on Linguistic and Cultural Diversity in Cyberspace*.

Let me say special thanks to the Russian Committee of the UNESCO Information for All Programme, the North-Eastern Federal University and the Interregional Library Cooperation Centre for this initiative.

I thank also the Government of the Republic of Sakha (Yakutia) for the warmth of their hospitality.

I must say it is a special pleasure to be in Yakutsk, in the heart of Siberia. This is a land of extremes – extreme weather and extreme beauty, and I think we will all experience the famous White Nights over the course of our stay. This is also a land of extreme wealth in terms of cultural and linguistic diversity, and this brings me to the theme of this Conference.

Our starting point is clear, and it has roots in UNESCO's 2001 *Universal Declaration on Cultural Diversity*:

As a source of exchange, innovation and creativity, cultural diversity is as necessary for humankind as biodiversity is for nature. In this sense, it is the common heritage of humanity and should be recognized and affirmed for the benefit of present and future generations.

This idea has never been so relevant – especially, in this *Year of Culture* of the Russian Federation, in the year when we celebrate the 25th anniversary of the World Wide Web, introduced by Sir Tim Berners-Lee in 1989.

This International Conference comes at the right time, when many societies are undergoing transformation, when the international community is shaping a new global development agenda to follow 2015.

This new agenda must do everything to safeguard cultural and linguistic diversity and harness its power for identities and belonging, for creativity and for dialogue.

This is essential to build the knowledge societies we need for the century ahead.

Societies today are more connected than ever. Information can be spread, received, and accessed with a click of a button. New technologies are revolutionising the way we communicate, create and share knowledge.

These trends are reshaping institutions – public and private – the economy, even our personal relations. They have spurred social transformation and advanced human development. They are also raising new challenges – challenges of access, of diversity of content, of multilingualism.

Languages are essential here. They are the foundation for human rights and dignity and the channel for communicating and sharing, for strengthening social cohesion and joint action.

Multilingualism is essential to the identities of people, to the strength of societies, to cultural diversity – we must do everything to preserve and strengthen this as a strength for all to share. This must start in cyberspace – which must provide a platform for all to share their heritage and culture, on the basis of human rights, and promote linguistic diversity.

The loss of linguistic and cultural diversity carries unmeasurable costs – for societies concerned, and, fundamentally, for all humanity. This loss would jeopardize meaningful development and it would hamper intercultural dialogue.

This is why this International Conference is so important.

In this spirit, let me thank once again the Russian Committee of the UNESCO Information for All Programme for its outstanding work.

Most of all, I wish to thank the participants, who come from many governments and multiple disciplines and from across the world, to share their insights.

I am confident this International Conference will set a new milestone in the commitment we share to promote linguistic and cultural diversity in cyberspace.

Greeting by Yegor Borisov,
Head of the Republic of Sakha (Yakutia)

Ladies and gentlemen,

I am happy to greet in Yakutia the participants and guests of the 3rd UNESCO conference on Linguistic and Cultural Diversity in Cyberspace.

Many ethnic cultures are assimilated and vanish under the impact of the current powerful globalization. Experts don't rule out the extinction of over a half of the present-day 7,000 languages within the lifetime of the several coming generations. Wise ethno-linguistic policies and the latest information technology are at least able to inhibit these trends, detrimental to the entire world.

We are glad that the Republic of Sakha (Yakutia) is, for a third time running, the venue of dialogue and initiatives on this essential matter when the united global information space is developing to attain a unique combination of ethnic cultures in their entire diversity. Public interest in the culture of the Russian North is growing. The appraisal of its socio-cultural role in interregional and international partnership provides a firm basis for the assessment of current global processes.

Of special importance in this context is the understanding of the present revolution in information and communications. Its fruit has a mighty impact on the public mentality as it changes the long-established cultural and moral norms and obliterates borders. This point mostly concerns the preservation and popularization of language culture.

Over 120 ethnic entities are represented in Sakha (Yakutia). Dominating our regional policy is the preservation and development of ethnic languages, cultures, customs and traditions. We are doing much to guarantee the promotion of Yakut, Russian and indigenous ethnic minorities' languages. We realize that it would be impossible to preserve linguistic and cultural diversity without ICT.

I am sure that this conference will come as another mighty impetus for comprehensive discussions of topical theoretical and practical problems pertaining to diversity in cyberspace, and will provide prerequisites for the further progress of this cause in Russia. I wish you efficient and fruitful work, ever new achievements, and luck in all your endeavours.

Greeting by Sergei Lavrov,
Minister of Foreign Affairs of the Russian Federation

I greet from the bottom of my heart the organizers and participants of the 3rd international conference on Linguistic and Cultural Diversity in Cyberspace.

The preservation of linguistic and cultural diversity acquires tremendous importance for sustainable development and in other spheres now that a new polycentric world is emerging and the meaning of civilizational identity is increasing apace. Certainly, these efforts cannot but spread to cyberspace in today's information society.

With many centuries' experience of interethnic and interreligious peaceful coexistence and cooperation, Russia actively promotes the linguistic and cultural diversity of the world. It has organized two international conferences on this theme within the frame of the UNESCO Information for All Programme. The final documents of these conferences – the Lena Declaration and the Yakutsk Appeal – propose how to implement the Recommendations for the World Summit on the Information Society, and advance the idea of a World Summit on Multilingualism, to be convened in 2017.

I am sure that the conference will contribute honourably to the profound and comprehensive analysis of problems on its agenda and will help its participants to learn better the affluent and original land of Yakutia.

I wish you every success in your fruitful work.

Greeting by Mikhail Seslavinsky, Head of the Federal Agency for Press and Mass Communications

Dear friends,

I am glad to greet you at the opening of the 3rd international conference on Linguistic and Cultural Diversity in Cyberspace. Indicatively, its venue is again Yakutsk, where speakers of two years ago expressed justified concern about the inevitable costs of information society's rapid development.

Progress brings us not only precious innovations but also many things we cannot put up with. Literally before our eyes languages are dying that were spoken quite recently by communities of people with their problems, joys and sorrows, with their unique culture.

We must spare no effort to rescue languages from extinction: they must actively develop in cyberspace to bring us all closer to the needs and interests of the world around us.

It is one of the noblest global goals to preserve healthy ethnic identity and the diversity of civilized languages because cultural diversity is an essential prerequisite of sustainable development, and of mutually respectful peaceful coexistence of individuals and nations.

We work to attain the latest standards of life based on universal human values. At the same time, we stay loyal to the best traditions of our nations and cultures. To build up the cultural potential and preserve international and interethnic peace and accord should be our most sacred goals. These goals are unattainable if we lack mutual understanding and fail to find the true solutions of our burning problems.

I wish the participants of the 3rd international conference on Linguistic and Cultural Diversity in Cyberspace every success, fruitful discussions and unforgettable impressions.

Greeting by Grigory Ivliev,
Secretary of State, Deputy Minister of Culture of the Russian Federation

Ladies and gentlemen,

I am happy to greet all the participants, guests and organizers of the 3rd international conference on Linguistic and Cultural Diversity in Cyberspace, which is opening in Yakutsk today.

Its theme is essential in a world where languages and cultures are extinguishing. Languages are the most precious treasure of the human race. They are the vehicles of historical experience and social and cultural traditions. They are the tools of self-expression and self-identification. Languages preserve the picture of the world, and this picture is unique as represented by every language.

The preservation of linguistic and cultural diversity is especially topical in multiethnic states. Russia is among them, with over a hundred indigenous ethnic entities, each preserving its original language and culture.

The 1st and 2nd international conferences on Linguistic and Cultural Diversity in Cyberspace gathered in Yakutsk under the UNESCO aegis in 2008 and 2011, respectively. They were organized by the Russian Committee of the Information for All Programme, the North-Eastern Federal University, the National Library of the Republic of Sakha (Yakutia), the Interregional Library Cooperation Centre, and the MAAYA – the World Network for Linguistic Diversity, with support of the governments of the Russian Federation and the Republic of Sakha (Yakutia). The conferences expanded and promoted professional contacts and helped to start personal friendships among leading international experts.

The agenda of such conferences is expanding with the geography of its representation. The first conference, in 2008, represented 15 countries, the second 30, and the present 50.These figures illustrate the growing influence and popularity of the conferences on Linguistic and Cultural Diversity in Cyberspace.

I wish you all fruitful discussions and new discoveries.

Greeting by Veniamin Kaganov,
Deputy Minister of Education and Science of the Russian Federation

Dear friends,

I am happy to greet the participants and organizers of the 3rd international conference "Linguistic and Cultural Diversity in Cyberspace", on behalf of the Ministry of Education and Science of the Russian Federation.

The protection and promotion of linguistic and cultural diversity is one of our national strategic priorities making the basis of our community's intellectual, moral, emotional and cultural development.

We cannot but notice the prominent role of UNESCO, under whose aegis this conference is working, in the formation and development of the global socio-cultural and socio-linguistic situation, and in addressing burning contemporary problems, particularly as cyberspace is developing apace and brings the danger of unifying languages and cultures.

Symbolically, Yakutsk is again hosting an international conference on Linguistic and Cultural Diversity in Cyberspace – the third this time. The Republic of Sakha (Yakutia) is populated by more than 120 ethnic entities and provides every condition to preserve cultural and linguistic diversity. It has a unique experience of how to promote cultural interpenetration and, at the same time, preserve cultural identity.

Of no smaller importance is the implementation in the educational systems of other parts of Russia of the practical patterns of nationalities policy elaborated in the Republic of Sakha (Yakutia), among them the methods and techniques of preserving linguistic diversity.

I am certain that the conference will promote contacts between its participants from many countries and allow them to exchange opinions and knowhow.

I wish you all successful work and fruitful decisions on the topical issues of promoting multilingualism in cyberspace, and every success in everyday life.

Greeting by Vyacheslav Nikonov, Education Committee Chair, State Duma, Federal Assembly of the Russian Federation

Ladies and gentlemen,

I greet from the bottom of my heart the organizers and participants of the 3rd international conference on Linguistic and Cultural Diversity in Cyberspace.

Your representative forum aims to discuss one of the most topical issues of the national and international policy of preserving and developing languages as cyberspace is rapidly extending. You will delve into the problems of preserving cultural identity on the Internet, the chances of internationalizing languages, the development of legislation, and the agencies to support linguistic diversity.

Command of the Russian language is useful, prestigious and even fashionable in the contemporary world. The Russian language is the basis of the multi-million Russian world. It is spoken in the best-known research centres and at essential economic and social forums. It was the first to be heard in the outer space, and is the second most important on the Internet. The prospects of its further international use depend on our joint efforts to preserve and develop it, and improve tuition in Russian in our country and abroad. To attain these goals is the principal prerequisite of making Russia intellectually richer and more competitive on a global scale.

I hope your debates will contribute considerably to the cause of preserving cultural and linguistic diversity in cyberspace and provide the basis of expert conclusions, practical proposals and legislative initiatives.

I wish you fruitful work, interesting discussions and useful professional contacts.

OPENING ADDRESSES

Getachew ENGIDA

UNESCO Deputy Director-General
(Paris, UNESCO)

Ladies and Gentlemen,

I wish to thank once again the Russian Committee of the UNESCO Information for All Programme, the North-Eastern Federal University and the Interregional Library Cooperation Centre for this initiative.

From the top, let me thank also the Government of the Sakha Republic (Yakutia).

This conference reflects the partnership UNESCO has developed with the Sakha Republic (Yakutia).

On 21 April, the UNESCO Director-General, Ms Irina Bokova, met with the President of Sakha (Yakutia), Excellency Mr Yegor Borisov, in Moscow – they signed a Joint Communique, renewing cooperation for quality education, in the sciences and environmental protection, in the social and human sciences, as well as in culture, and in communication and information. The Joint Communique built on solid grounds.

This Wednesday, I know a visit is organised to the *Lena Pillars Nature Park* – these spectacular rock pillars, stretching along the banks of the Lena River, are inscribed on the UNESCO's World Heritage List.

In 2008, the *Yakut Heroic Epos Olonkho* was inscribed in the UNESCO Representative List of the Intangible Cultural Heritage of Humanity. Weaving narration and song together, this epic reflects the world of knowledge accumulated by the Yakut people over the centuries.

On a personal note, I recall well the *Days of Yakutia* at UNESCO in Paris, on 21 March, 2012 – when I was honoured to welcome Excellency, Galina Dantchikova, Prime Minister of the Government of the Sakha Republic.

This partnership has developed in the framework of deep cooperation between UNESCO and the Russian Federation.

In Moscow, last April, the UNESCO Director-General attended the ceremony to celebrate 60 years of membership of the Russian Federation in UNESCO, with Excellency Sergey Lavrov, Minister of Foreign Affairs and President of the Commission of the Russian Federation for UNESCO.

This was an opportunity to highlight 60 years of action for the ideals of UNESCO, for the values we share. The values of equality, dignity and mutual respect. The values of dialogue and cooperation.

These same values have brought all of us to Yakutsk today, for this *3rd International Conference on Linguistic and Cultural Diversity in Cyberspace*.

Cultural and linguistic diversity stands at the heart of the UNESCO Constitution, which calls for building the defences of peace in the minds of women and men, through the free flow of ideas by word and image.

In my opening remarks, I cited UNESCO's 2001 *Universal Declaration on Cultural Diversity* – let me quote the Declaration again:

The defence of cultural diversity is an ethical imperative, inseparable from respect for human dignity. It implies a commitment to human rights and fundamental freedoms, in particular the rights of persons belonging to minorities and those of indigenous peoples. No one may invoke cultural diversity to infringe upon human rights guaranteed by international law, nor to limit their scope.

On this foundation, UNESCO takes a multi-disciplinary approach to safeguarding and promoting cultural and linguistic diversity.

This starts with work to support multilingual education and to promote the use of mother tongues – this is essential in increasingly multicultural societies. Education today must be about learning to live together as well as learning to know, to do and to be.

Our work includes support to countries across the world to implement the UNESCO Culture Conventions, to safeguard humanity's shared cultural heritage. It involves promoting local content and linguistic diversity on the Internet.

Languages lie at the heart of UNESCO's action.

Languages provide the lens through which the world is understood and the material through which it is voiced. They express the values we share and give shape to ideas, linking the past with the future.

It is through language that we make sense of the world and that we can transform it for the better.

Multilingualism is important, because it opens opportunities for mutual understanding and cooperation, because it creates a plural linguistic space, which allows the wealth of diversity to put in common. Multilingualism is a force for inclusion and social cohesion – it is also a foundation for global citizenship.

Promoting global citizenship is a key goal of the United Nations Secretary-General's *Global Education First Initiative*, which UNESCO is steering forward.

Nelson Mandela once said: *"If you talk to a man in a language he understands, that goes to his head. If you talk to him in his language, that goes to his heart."*

In a world of rising diversity, language ability is vital for intercultural understanding. This is why the loss of any language is a loss for all humanity. It is a loss for human memory, for shared knowledge, for the linguistic and cultural diversity that is our common heritage, and a cornerstone for peace and reconciliation.

And yet, an estimated 50 percent of the world's 6,700 spoken languages are in danger of disappearing, and many more face the threat of declining influence.

This is one of the challenges we have met to address.

The digital revolution offers a number of answers – provided we harness its power to preserve and promote diversity.

New information and communication technologies are opening new frontiers for innovation, creativity and development. The Internet is widening opportunities for cultural expression and dissemination. The lowering of the cost of digital technology or equipment, along with lower Internet access costs and the introduction of Internationalised Domain Names, provide unprecedented opportunity for people to access, produce and share content globally.

The Internet must be central to all efforts to promote linguistic and cultural diversity – and this must proceed on the basis of human rights, which must be respected both offline and online, in accordance with international human rights obligations and standards, as well as UNESCO decisions.

Digital local content is proliferating, thanks to growth in developing countries.

Cheaper and faster smartphones and tablet computers are bringing Internet access to more people in more places. Every year, new languages are becoming available on these platforms – this allows those who speak endangered languages to create content, and speakers of every language to share in the language of their choice.

At the same time, opportunities are accompanied by challenges.

The challenge of access, as not everyone can take advantage of technological progress. Even where there is broad access to the Internet and other ICTs, this does not guarantee that everybody is able to participate, contribute and benefit equally.

The digital divide continues to deepen.

Less than five percent of world languages are used online. The Internet and ICTs raise some tough concerns for governments, for professional communities, for users of minority and lesser-used languages.

More and more users develop web content in English, a lingua franca that is neither their mother tongue nor a national or regional language – this means that the less content available in a particular language, the higher its risk of digital extinction, as users and developers migrate away.

Challenges include limited resources to implement policies for multilingualism.

Internet services in many States remain costly, largely unavailable, and slow. The development of local technical skills and expertise is progressing too slowly. The low level of digital literacy and the undeveloped info- and infra-structures are creating barriers for marginalized groups to access information and knowledge on the Internet – I would highlight here the particular needs of persons with disabilities.

In addition, a host of ethical questions is being raised – we need to ensure that universal values and fundamental rights are promoted and respected in cyberspace.

These are just a few of the challenges we must address, to ensure the digital divide does not hold back entire societies from sustainable development, from the information and the means of communication necessary for health and education, from opportunities to take part in cultural, political and economic development.

Everyone should have access to a multilingual Internet and content. But this will not happen on its own.

We need to allocate greater resources, to provide tools and to take concrete measures to support all languages on the internet.

Ladies and Gentlemen,

This vision guides all of UNESCO's work to build knowledge societies that are inclusive, pluralistic, equitable, diverse, open and participatory.

Our action starts at the normative level – with the *Recommendation concerning the Promotion and Use of Multilingualism and Universal Access to Cyberspace*, adopted by Member States in 2003. This Recommendation provides clear guidance on steps to be taken to advance multilingualism in cyberspace.

Just this month, we invited all Member States to report for the third time on progress towards the implementation of the recommendation, to develop

a report that will be submitted to Member States though the UNESCO Governing Bodies.

We are working at the global level to promote multilingualism on the Internet. This is an important part of UNESCO's contribution to the World Summit on the Information Society – where we facilitate implementation of the Action Line C8, *Cultural and Linguistic Diversity* – as well as our cooperation with the Internet Governance Forum.

The same spirit guides our vice-co-chairmanship of the *Broadband Commission for Digital Development*, set up by UNESCO and the ITU – to promote global, accessible and inclusive broadband roll-out for sustainable development, where we support the Working Group on Multilingualism.

We are also working with the *Internet Corporation for Assigned Names and Numbers* – ICANN – with whom we signed a cooperation agreement in 2009, to promote multilingualism.

With EURid, we are monitoring the deployment of Internationalised Domain Names, through global reports, to enhance online linguistic diversity and access to multilingual content.

We are also active on the ground with Member States.

UNESCO has supported States in Latin America, training decision-makers to implement recommended policy measures in these areas. Similar activities are planned this year in Central America, on issues related to indigenous peoples and the Internet.

UNESCO's *Atlas of the World's Languages in Danger* remains a flagship of our work, and we will scale up the online platform, under the leadership of its Communication and Information Sector.

We continue to lead research, to understand trends and craft better policies in response.

With the OECD and the Internet Society, we are embarked on a second study on the relationship between local content, Internet development and access prices. With ICANN and the Internet Society, we are working to develop language tools, such as the glossary on Internet governance terms for Arabic speakers.

Ladies and Gentlemen,

The stakes are high, because languages do not only express the world – they shape it.

Language is the bridge between ideas and action – it is an essential part of what we at UNESCO call a new humanism, rooted in respect for human dignity, fundamental rights and the diversity of cultures.

This is why we must do everything to promote cultural and linguistic diversity in cyberspace.

I believe there is a Yakut proverb that says: *"The blacksmith and the shaman are of the same nest."* The truth is, with language, we are all both blacksmiths and shamans, forging new forms of meaning, creating new materials for understanding, through words, through shared expressions.

We must protect this power for all.

Evgenia MIKHAILOVA

*Rector, North-Eastern Federal University
(Yakutsk, Russian Federation)*

Ladies and gentlemen,

Ethnic identity and unity are manifest in ethnic languages and cultures. Yakut literary classic Alexei Kulakovsky, the founder of Yakut artistic writing, said that the neighbouring big and small ethnic entities did not develop evenly, and warned that territorial proximity could lead ethnic minorities to utter assimilation and, in the final analysis, extinction.

Yakutia's multi-ethnicity is more than its specific feature – it's the source of Yakutia's wealth and spiritual strength. That is why the preservation and promotion of linguistic and cultural diversity is one of the principal goals of its state policy. We are building an open society that cherishes its linguistic diversity and encourages respectful interest in other peoples' languages and cultures while implanting love of one's mother tongue and native culture.

Allow me to greet you in the ancient land of the Olonkho during the **Yssyakh Ethnic Festival**, and wish you wellbeing and happiness. The Yssyakh is a traditional festival of the Sakha people, which manifests the beginning of summer and people's creative unity.

The older people and cultural historians say that one must necessarily receive a blessing during the festival from the *algyschyt* priests, whose sacred rites strengthen creative drive and assurance.

Cultures are levelled out before our very eyes. Unification is sweeping out their diversity and brings closer the doom of minor languages. Forecasts say that up to 90% of the present-day 7,000 languages will be utterly forgotten by the end of the 21st century. We would like to hope that this forecast is wrong. Supporting this hope are such events as this international conference on the preservation and development of languages in cyberspace, which focus the search for tools and patterns of linguistic development. Such events are also centres of inspired persuading.

As one of its organizers, the North-Eastern Federal University regards the conference not as a mere platform to discuss burning problems but as a global expert forum on practical measures to preserve and develop cultures and languages – a mission worthy of a federal university.

A classic said once that to invent the future was the best way of forecasting it. The Lena Resolution was adopted to summarize the first forum. The second brought the Yakutsk Call for Action, a Roadmap towards the World Summit

on Multilingualism. We make it a point to implement all recommendations, and ever more experts from many countries join the cause with every passing year. I am sure that Yakutsk has proved its value as the venue of profound expert debates and the choice of effective measures to promote minor languages.

In 2011, the North-Eastern and Siberian federal universities launched together a foresight study of the development of northern areas and their indigenous population up to 2050. Extensive use of expertise sets foresight studies apart from the more conventional prognostication. As expert studies show, the next decade or two will bring sweeping social, economic and cultural change to the Republic of Sakha (Yakutia). It will be a two-fold change: on the one hand, it will boost economic progress and spectacularly improve the quality of life while, on the other hand, it may critically change the indigenous peoples' life. The initial stages of our studies have brought information allowing assess the pace at which Northern ethnic cultures and languages have been dying out since 1950, to forecast their development up to 2050, and recommend practical measures for the preservation and reproduction of the languages and cultures of the Russian north-east.

It is of principled importance that foresight studies include the analysis of the future's hazards and opportunities, i.e., the study of positive and negative trends and crisis prognostication. This is the greatest merit of foresight studies, which show what to do considering tentative hazards instead of making goody-goody pictures of the future seen through rosy specs.

As foresight studies show, folk festivals are important to all age groups and so come out as connecting links between generations, as half of our respondents said. At the same time, opinion polls warn about the risk of the **generation gap widening** with the use of ethnic customs and codes of conduct. The younger generation holds folk mentality, legends, music, athletic games and medicine in far smaller esteem than the older.

Everyday use of ethnic languages is an essential factor of its survival. When even people fluent in their native language prefer to discuss their everyday affairs in another, the mother tongue is gradually ousted into the background to be used only on specific occasions. Our poll shows major generation differences in everyday use of ethnic languages: the younger the respondents, the rarer they discuss home and personal affairs in their native language (85% in the age group of over 60, 80% in the 50–59 group, 79% in 40–49, 76% in 30–39, and 68% in 20–29).

The respondents point out the insufficient influence of educational institutions as instruments of promoting ethnic culture – only 8% mentioned them at all, which means that the North-Eastern Federal University and other

educational establishments should enhance their efforts for its reproduction and development.

Expert knowledge helps to assess the Yakutian population's present and forecast the future, particularly the developmental trends of languages and culture.

The increasing use of ICT has a dual impact on linguistic diversity: on the one hand, it dooms languages to premature oblivion – only 7% of presently existing languages occur in the Internet, while, on the other hand, ICT provides new tools to preserve and revive minority languages. It is up to us to decide whether our native languages, and our mentality connected with them, have the right to live on.

English language domination gains momentum, promoting the United States' and the entire West's political, economic, academic and cultural interests.[1]

The first international conference on Linguistic and Cultural Diversity in Cyberspace was convened **in 2008**, which the United Nations proclaimed International Year of Languages. The conference was held within the frame of the UNESCO Information for All Programme, under the auspices of UNESCO and the Government of the Russian Federation, and was supported by the Russian Federation ministries of Culture and Foreign Affairs, and the Government of the Republic of Sakha (Yakutia). The first major international forum dedicated to the burning issue contributed spectacularly to International Year of Languages. With 15 nations represented, the conference demonstrated a positive image of Russia as a multiethnic country with an effective and comprehensive policy of indigenous language and culture promotion and development – suffice to say that only three languages became extinct in Russia over the previous 300 years, as compared to over 80 that got completely out of use in the United States. The first conference brought plans of practical action and endorsed the **Lena Resolution**.

National representation doubled to 30 countries at the second conference **in 2011 г**. The conference discussed relevant experience, summarized research and adopted the **Yakutsk Call for Action**, a Roadmap towards the World Summit on Multilingualism.

We are glad that the world approves Yakutia's steps to promote linguistic and cultural diversity and heritage. We are glad that a community of top-notch Western experts has gathered round us to become Russia's friends whom we address for help and support.

---

[1] *Russia is assessing Chinese IT companies' proposal to establish new telecommunication corridors in the Far East so as to bypass US servers in Russian-Chinese information exchange*

The conferences upgraded our cultural, research and educational efforts, gave them methodological support and enriched them with horizontal contacts in Russia and all over the world. Our libraries, archives, universities and research institutes now communicate not only between themselves but also with software manufacturers to improve their bilingual websites and other resources.

Our joint initiative for a third conference has found extensive support of the Government of the Republic of Sakha (Yakutia), the Federal Agency for Press and Mass Communications, the ministries of Culture and Foreign Affairs of the Russian Federation, and the Commission of the Russian Federation for UNESCO. It gathers under UNESCO auspices. The Russian delegation announced the upcoming conference during the UNESCO General Conference of November 2013, and sent invitations to all national commissions for UNESCO, all countries' ambassadors to UNESCO, and top-notch relevant experts the world over.

The North-Eastern Federal University is aware of the necessity to preserve languages as the principal cultural aspect at all levels. A majority of our magisterial programmes are implemented at the Institute of Languages and Cultures of the Peoples of the Russian North-East, at the university Department of Philology, the Institute of Foreign Philology and Regional Studies, and the Institute of Mathematics and Information Technology.

The university is launching another 38 magisterial programmes in 2015, some of them on cultural and folklore studies. The university maintains international and interethnic contacts in and outside Russia, and will implement a part of the programmes online, in cooperation with other universities and research institutes. For instance, we cooperate with Kazan Federal University in philology, with Universit de Versailles Saint-Quentin-en-Yvelines in cultural heritage, environment and tourism, and St Petersburg State University of Culture and Arts in cultural history.

Active university work has found reflection in national and international ranking. It is on the top 200 list of 6,000 BRICS countries' universities alongside another 52 Russian-based universities. It ranked 38th out of 1,500 on the 2013 national university list. All this proves that the North-Eastern Federal University is among Russia's leading educational institutions.

Contemporary education both reflects and supports cultural and linguistic diversity. Every stage of social development demands reappraisal and readjustment of educational goals, particularly in the preservation and development of ethnic languages and cultures. The new generation of our university experts is growing in an atmosphere of well-wishing respect as they learn to think and work in many fields using several languages.

Education demands reform to provide quality that would guarantee students' active life and employment in a globalizing world. It should, however, retain its specifics rooted in the linguistic situation in its region – mosaic diversity, in our instance.

We want to develop the Russian language because it is not only the national treasure of Russia and ethnic Russians abroad – it is a world treasure as well. It is essential to preserve and extend Russian cultural presence in other countries. We will continue supporting Russian language studies in overseas universities, and will assist Russian language and literature chairs there.

I have just returned from an East Asian Slavic scholars' conference in South Korea, where I made a plenary report in support of Russian as a language of international communication. The conference approved the North-Eastern Federal University's appeal to launch a project for international comparative studies on the preservation of linguistic diversity in many countries of the world. Professor Kang Duk Soo of the Hankuk University of Foreign Studies (South Korea), professor emeritus of the North-Eastern Federal University, has kindly agreed to lead the project.

We are working actively to preserve and develop the **Yakut language and the languages of the Northern indigenous ethnic minorities**. Everyone is welcome to Yakutia's official language classes. The university provides higher education in the Yakut language at its Institute of Languages and Cultures of the Peoples of the Russian North-East. Through research and public educational activities in the study and preservation of artistic and intangible cultural heritage of the Russian North-East, the institute seeks to integrate research with practical education, guarantee the dynamic development of the languages, literature and culture of its indigenous peoples – the Yakut, Evenk, Yukagir, Dolgan, Chukchi, Koryak and Aleut – and develop bilingual education.

As part of its development programme for 2010–2019, approved by the federal government, the university is implementing its **programme for the preservation and development of the Northern indigenous ethnic minorities' languages and cultures in cyberspace and digital recording**. The preceding four years saw ambitious work done on the basis of the university New Information Technology Centre, with four major ethnological expeditions to indigenous peoples' areas of compact settlement. The university has produced 17 unprecedented multimedia educational complexes on the indigenous peoples' languages, culture and folklore, and established the **www.arctic-megapedia.ru** website. It carries information on the languages and cultures of the Russian North-East's indigenous ethnic minorities, and in forming an archive of full text documents and audio and

video resources. Information is available in Russian, English and the official minority languages.The project can be extended to the entire Russian area of Northern and Siberian ethnic minorities' settlement. The university has elaborated the multimedia digital archive of Northern and Siberian ethnic minorities – specialized software allowing to place full text information sources, archive documents and photos, and audio and video files.

The centre is also engaged in a comprehensive project to track down and study specific symbols/letters of Northern and Siberian minority language fonts missing in computer operating systems. Over forty have been found for today. The work must go on with adequate government funding to include them in UNICODE.

The comprehensive assessment of the North-Eastern Federal University's role and potential in regional development shows that the preservation and development of Northern peoples' languages and culture is a new and essential area of university work. It comprises professional education, research, public education in history and culture, multilingual and multicultural education, social engineering and cultural policy.

The North-Eastern Federal University is called upon to become a **strategic centre** for the formation of cultural, research and educational environment of Russia's North-East – a centre resting on ethnic cultural values, and a stronghold of lasting cultural partnership. Of special global cultural significance is innovative university research to implement an academic information system to preserve and disseminate the Olonkho Yakut heroic epic. An Olonkho research institute, a special television channel, and an information portal have been established.

To implement the Lena Resolution, a decision was made to establish a centre to advance multilingualism within the university cyberspace. Due to the evaluation standards of the survival of languages in cyberspace, elaborated by its staff, the centre can now contribute to one of the principal causes of this conference – the distribution of roles, functions and responsibilities for education and the preservation of cultural diversity. These standards help to assess the kind and amount of government assistance necessary to preserve a particular language.

We have gathered here today to show to the world that multilingualism is a norm of the contemporary community.

I wish all conference participants fruitful work for our common cause of the preservation and development of world languages.

Evgeny KUZMIN
Vice-Chair, Intergovernmental Council,
UNESCO Information for All Programme (IFAP);
Chair, Russian National IFAP Committee;
President, Interregional Library Cooperation Centre
(Moscow, Russian Federation)

Multilingualism in Russia

Introduction

Ladies and gentlemen,

This is a third international conference I am organizing in Russia on the preservation of linguistic and cultural diversity in reality and the development of multilingualism in cyberspace. However, I dare only now to make a comprehensive coverage of the situation of multilingualism in Russia. I mentioned this theme in passing during my presentations at the previous conferences. Now I make it the sole theme of my address.

Russia is a multilingual country though this fact is hardly known outside it. At the time of the Soviet Union, many people in the world knew or guessed that such a huge country should be multiethnic. However, very few truly realized the fact, as I have seen recently. It is dawning upon them now that Russia, which accounts for a half of the former Soviet population, is also multiethnic. At any rate, almost all my educated foreign colleagues, including Europeans, were greatly surprised when I told them that Russians are not only ethnic Russians.

Every European understands that Russia, as any other major country, should shelter many immigrants, and they realize that there might be many diasporas in Russia since the time of the Russian empire and later the Soviet Union. But they are really stunned when I tell them that there are another hundred indigenous ethnic entities in Russia. By "indigenous" I mean entities historically formed within Russia's present borders or ones whose majority has lived here for several centuries and who have no statehood and large populated areas outside Russia.

What is really stunning is that people are unaware of this even in Russia. To be honest, I myself realized vast Russian multilingualism quite recently, in 2006, after I took up multilingualism in cyberspace professionally on request of the Commission of the Russian Federation for UNESCO. Everyone in Russia certainly knows that it's a multiethnic country – but when I ask my Russian

friends, even university people, how many indigenous ethnic entities there are in Russia and how many languages they speak, they are sent into consternation. Only few give precise answers. More than that, when President Vladimir Putin said proudly a year ago that Russia had retained and was developing the languages of almost all its indigenous peoples, he added that he had learned it quite recently.

Our educational system is rather good still, as the whole world knows. Russians study history, geography and the ABC of social science since childhood but never pay attention to the survival of multilingualism, however remarkable and praiseworthy it might be. I think it is a huge error. We have grown accustomed to taking pride in the sublime Russian past, in Russia's achievements in the arts, culture and research, in our space effort, etc., but we have, I think, only recently opened our eyes to our breathtaking cultural diversity that goes hand in hand with our vast cultural heritage. We are only learning to take pride in this diversity, to which we paid little attention in the past, taking it for granted.

Now, we are traveling more than ever before, and have the opportunity to compare Russia to other countries. That is why we better understand our own country and value it higher. When we hear numerous appeals to other nations at the political level worldwide – appeals to tolerance, persuading the world to reckon with ethnic minorities' rights, we grow to realize that Russia is truly tolerant to them. More than that, throughout its history it has consciously and purposefully protected their cultural identity and promoted their languages not in word but in deed.

Books and press outlets are published in almost all indigenous languages in Russia. They are tuition languages, at least at primary school. They are television and radio broadcasting languages. Internet information resources in these languages are developing. All languages are studied and documented painstakingly. All are treated as precious things. They matter tremendously to the Russian state and the Russian public because we have long ceased to qualify people as first and second rate according to ethnicity. All are brothers to us. In the Soviet times, parents and schoolteachers taught me to treat all as brothers. Georgians were my brothers, just as Azerbaijanis, Kazakhs, Letts, Lithuanians, and others. More than that, we really regarded Poles, Czechs, Hungarians and all other socialist countries' people as brothers, to say nothing of Ukrainians and Belarussians.

I think it was a real breakthrough and I don't think any other major multilingual country has achieved as much.

Russia is not only one of the most multiethnic and multilingual countries in the world but also one of the most polyreligious. Not only Christianity, Islam

and Judaism but also paganism has firm historical roots here. There are also two Buddhist ethnic communities – Buryats and Kalmyks. When you ask a European whether there is a Buddhist ethnic entity in Europe, the answer is usually "no". That's wrong: there are Kalmyks, the offsprings of Mongolian tribes who came in the 16th into early 17th century from Central Asia to the lower reaches of the Volga and the north Caspian coast. They have their own statehood in the Republic of Kalmykia within the Russian Federation.

Russia respects and cherishes ethnic languages because it respects all its indigenous peoples and treats them as brothers.

Let us analyze Russia's ethnic composition before we go on talking about languages spoken in this country.

The Ethnic Composition of the Russian Federation

The Russian population made **142,856,536**, according to the 2010 census. They belonged to 245 ethnic entities, 100 of them indigenous.

Table 1 specifies the numerical strength of the 30 largest ethnic entities. The names of entities whose representatives have been living in Russia for a long time while having states or major populated areas outside Russia are italicized.

Table 1

No	Entity	Strength, persons	Portion of entire Russian population
1	Russian	111,016,896	77.71%
2	Tatar	5,310,649	3.72%
3	*Ukrainian*	1,927,988	1.35%
4	Bashkir	1,584,554	1.11%
5	Chuvash	1,435,872	1.01%
6	Chechen	1,431,360	1.00%
7	*Armenian*	1,182,388	0.83%
8	Avar	912,090	0.64%
9	Mordovian	744,237	0.52%
10	*Kazakh*	647,732	0.45%
11	*Azerbaijani*	603,070	0.42%

No	Entity	Strength, persons	Portion of entire Russian population
12	Dargin	589,386	0.41%
13	Udmurt	552,299	0.39%
14	Mari	547,605	0.38%
15	Osset	528,515	0.37%
16	*Belarussian*	521,443	0.37%
17	Kabardian	516,826	0.36%
18	Kumyk	503,060	0.35%
19	Yakut	478,085	0.34%
20	Lezgin	473,722	0.33%
21	Buryat	461,389	0.32%
22	Ingush	444,833	0.31%
23	*German*	394,138	0.28%
24	*Uzbek*	289,862	0.20%
25	Tuva	263,934	0.19%
26	Komi	228,235	0.16%
27	Karachai	218,403	0.15%
28	*Gypsy*	204,958	0.14%
29	*Tajik*	200,303	0.14%
30	Kalmyk	183,372	0.13%

Russian Nationals and Ethnic Russians

When we talk about the ethnic composition of the Russian Federation in English, we ought to distinguish two different phenomena: 1) ethnic Russians and 2) all Russian nationals (the entire population of Russia). The English language, literature and media outlets most often use one word, "Russian", for both. Laymen, i.e., not experts on Russia, most often understand it as ethnic Russians, referring at once to ethnicity and nationality.

The present-day Russian vocabulary has two categories to designate the two phenomena and distinguish between them: 1) *russkie*, pronounced as rouss-ki-je – mostly meaning ethnic Russians and 2) *rossiyane*, pronounced as ros- see- ya-neh, referring to all Russian citizens (the term is unambiguous, concerning only citizenship but by no means ethnicity).

I often visited America and talked to American men and women about their ethnic identity and background. When I heard that their grandparents were Italian immigrants and the parents of his/her spouse were also of Italian ancestry, I said every time: "So you're not American! You are Italians resident in the United States," receiving every time a heated rebuff: "We're American! It's our ancestors who were Italian!"

Everyone who lives in America is American. In Russia, things are quite different.

When they are in Russia or communicate in Russian, Tatars, Yakuts, Udmurts or members of any other indigenous ethnic entity never say they are Russian when asked in Russian about their ethnicity. They say: "We are Tatar/Yakut/ Udmurt," etc. When they get together, they never say: "*We are russkie,*" but "*We are rossiyane.*" But when abroad, especially in an English-speaking country, or during a talk in English, they most probably pose as Russians not "Rossiyane" not to go into detail and to avoid more questions.

The State Structure of Russia and Ethnic/State Autonomies

It is essential to see that the state structure and administrative territorial system can of themselves promote the preservation and development of minority languages or intensify their marginalization. A unitary multiethnic state strengthens and paces up cultural unification and ousts all languages except the official ones into the background. A federation, on the contrary, slows down the extinction of languages and is able to promote their development.

The Russian Federation possesses a sophisticated structure with 85 constituent entities – 46 regions, 9 territories, 22 republics, 4 autonomous areas, an autonomous region, 3 federal cities.

A region, or *oblast*, is an administrative territorial entity not merely dominated by ethnic Russians. It has no localities densely inhabited by other ethnic entities or, at least, they account for less than 1% of the population.

A territory, or *krai*, is a major administrative territorial entity that includes autonomous areas of ethnic minorities' compact settlement.

Republics are constituent entities populated by numerically comparable communities of Russians and other ethnic entities, large enough according to the standards of the Russian Federation. Republics are named after such entities. For instance, the Republic of Tatarstan owes its name to Tatars populating the area for a long time; the Republic of Buryatia is named after Buryats, etc. The constituent republics of the Russian Federation have their own constitutions and possess greater independence from the federal centre than territories, regions and autonomous areas.

Most of major or medium-size indigenous ethnic entities enjoy autonomy. Autonomies are constituent entities of the Russian Federation.

Turkic autonomies:

- Republic of Tuva — the Tuvinian make 77% of the population;
- Republic of Chuvashia — the Chuvash and Tatar account for 70% of the population;
- Republic of Bashkortostan — the Bashkir, Tatar and Chuvash, 57%;
- Republic of Tatarstan — 56%, Tatar and Chuvash;
- Republic of Sakha (Yakutia) — 47%, Yakut;
- Republic of Karachai-Circassia — 44.3%, Karachai and Nogai;
- Republic of Altai — 40%, Altaian;
- Republic of Dagestan — 20.6%, Kymyk, Nogai and Azerbaijani;
- Republic of Kabarda-Balkaria — 14.8%, Balkar, Tatar and Turks;
- Republic of Khakassia — 12%, Khakass.

Finnish-Ugrian autonomies:

- Republic of Mari El — Mari, 43.9%;
- Republic of Mordovia — Moksha and Erzya, 40%;
- Republic of Udmurtia — 28%, Udmurt;
- Republic of Komi — 23.7%, Komi;
- Nentsi Autonomous Area – 18%, Nentsi;
- Republic of Karelia — 9.3%, Karel, Finnish and Vepsian;
- Yamal-Nentsi Autonomous Area — 5.9%, Nentsi;
- Khanty-Mansi Autonomous Area — 1.9%, Khanty and Mansi.

As was said above, the Russian Federation also includes the following constituent entities: the Republic of Kalmykia, the Republic of Buryatia and the Chukchi Autonomous Area.

The Kamchatka Territory includes the Koryak Autonomous Area, the Krasnoyarsk Territory the Evenki Autonomous Area, the Trans-Baikal Territory the Ust-Ordynsky and the Ust-Buryat autonomous areas, and the Perm Territory the Komi-Permyak Autonomous Area.

Consistent efforts are made throughout Russia to preserve cultural and linguistic diversity. Constituent republics are the sites of the largest-scale and most active efforts to promote multilingualism and enhance the status of titular ethnic groups' languages in reality and cyberspace alike.

The State Languages of the Constituent Republics of the Russian Federation

According to a universal rule, Russian and the language of the titular ethnic group, to which a republic owes its name, are recognized as the state languages of the republic even when this group is an ethnic minority in its republic. Thus, the Bashkir make mere 30% of the four million population of the Republic of Bashkortostan, one of the largest constituent entities of the Russian Federation, while Russians account for 43.6%.

In some republics, two or more languages spoken there have the official status. For instance, Kabardian-Circassian and Karachai-Balkar are state languages, apart from Russian, in Kabarda-Balkaria, and Moksha and Erzya in Mordovia.

The Republic of Sakha (Yakutia) is among the unique places of the world for the survival of languages. Yakut, the language of the small titular ethnic entity, is developing there while the Yakut people support and promote the languages of the Northern indigenous ethnic minorities. Even, Evenki, Yukagir, Dolgan and Chukchi have the status of official languages in the republic, however few people speak them.

Of special interest is the situation – unique in certain respects – in the Republic of Dagestan in the North Caucasus. It has more than 120 ethnic entities but no officially recognized titular ethnic group, whose political attributes belong to 14 entities. Their languages belong to three language families: the Dagestani-Nakh branch of the Iberian-Caucasian language family, the Turkic group of the Altai language family, and the Indo-European language family. The Constitution of the Republic of Dagestan says: "Russian and the languages of the peoples of Dagestan are the state languages of the Republic of Dagestan," without enumerating the Dagestani peoples or languages – not through negligence but due to the extreme importance and sensibility of those matters in the republic. As certain Dagestani authors point out, the local practice has shown more than once that whatever attempt to make a legally binding closed list of ethnic entities and languages inevitably arouses a storm of protest and disputes that **defy settlement in principle**. The language situation in Dagestan is so complicated also because we do not know to this day how many languages there are presently in the

republic. References are usually made to sixty independent verbal languages. "Every mountain has a people of its own, and each speaks its own language," a local joke says.

The establishment of state languages does not mean that the other languages spoken in Russia are doomed. On the contrary, every ethnic entity has the guaranteed right of preserving, studying and developing its native language. Tatarstan, for instance, does much to preserve the culture and language of the local Bashkir, Udmurt and Chuvash, while Chuvashia promotes Tatar and Bashkir culture and Bashkortostan does the same for Tatars, Udmurts and Chuvashes. These three republics with Turkic languages predominant coexist peacefully with Udmurtia and Mordovia, with their Finnish-Ugrian population, which do much to preserve the Tatar, Bashkir and Chuvash languages.

Russia is also unique for the number of state and official languages of ethnic republics – the total approaches forty.

The Ethno-Linguistic Composition of the Russian Population

Russian is the official language of Russia, used almost everywhere in the country for interethnic contacts. It is the most widespread of all languages used in this country – a language renowned for its literature and scientific works; the language of a universally respected educational system. It is also a countrywide language of official paperwork. Russian largely retains its functions in the former Soviet republics, now independent states.

More than 127 million people regard Russian as their native language. A majority of other ethnic communities have its fluent command. Many know Russian better than their own mother tongue, and some even better than many ethnic Russians. 13 million of non-Russians regard Russian as their native language. Some don't know their mother tongue at all. They are especially numerous among people who were born in a big city and live there now.

Table 2 distributes the Russian population into language families and groups, which consist of indigenous peoples whose majority lives in Russia and who have no statehood and no large diasporas outside Russia (after 2010 census statistics).

Table 2

Language family	Number of speakers, 2010	%
INDO-EUROPEAN FAMILY	116,618,315	81.633%
• Slavic group	113,545,778	79.482%
• Iranian group	807,002	0.565%
ALTAI FAMILY	12,737,769	8.916%
• Turkic group	12,011,825	8.408%
• Mongolian group	647,761	0.453%
• Tungus-Manchurian group	78,183	0.055%
NORTH CAUCASIAN FAMILY	5,058,304	3.541%
• Nakh-Dagestani group	4,284,987	3.000%
• Abkhaz-Adyg group	773,317	0.541%
URAL FAMILY	2,371,398	1.660%
• Finnish-Ugrian group	2,322,020	1.625%
• Samoyed group	49,378	0.035%
CHUKCHI-KAMCHATKA FAMILY	28,985	0.020%
NIVKH (isolated language)	4,652	0.003%
YUKAGIR FAMILY	2,605	0.002%
ESKIMO-ALEUTIAN FAMILY	1,738	0.001%
YENISEI FAMILY	1,219	0.001%

Russia's most widespread languages beside Russian are Tatar (5.35 million speakers), Bashkir (1.38 million), Chechen and Chuvash (1.33 million each).

There are another nine languages with the number of speakers varying from 400,000 to a million: Avar (785,000), Kabardian-Circassian (588,000), Dargin (504,000), Osset (494,000), Udmurt (464,000), Kumyk (458,000), Yakut (456,000), Mari (451,000) and Ingush (405,000).

Another 15 indigenous languages are spoken by 50,000 to 400,000: Lezghian (397,000), Buryat (369,000), Karachai-Balkar (303,000), Tuva (243,000), Komi (217,000), Gypsy (167,000), Kalmyk (154,000), Lak (153,000), Adyghei

(129,000), Tabasaran (128,000), Komi-Perm (94,000), Nogai (90,000), Altai (66,000), Karel (53,000) and Khakass (52,000).

All languages spoken in Russia except Russian are minority languages and are affected by marginalization to varying extents because members of ethnic minorities who have no fluent command of Russian cannot aspire to a good career and self-fulfilment in the intellectual sphere.

Russia's Endangered Languages

More than a third of languages spoken in Russia are endangered or extinguishing. The situation is the worst for the languages of ethnic minorities less than 50,000-strong, mainly belonging to the indigenous population of the Far North, Siberia and the Far East:

- 25,000–50,000 speakers – Nentsi (41,302), Evenki (35,527) and Khanty (28,678);

- 10,000–25,000 – Even (19,071), Chukchi (15,767), Shor (13,975), Nanai (12,160) and Mansi (11,432);

- 1,000–10,000 – Koryak (8,743), Vepsian (8,240), Dolgan (7,261), Nivkh (5,162), Todjin-Tuva (4,442), Selkup (4,249), Itelmen (3,180), Kumandin (3,114), Ulchi (2,913), Soyot (2,769), Teleut (2,650), Telengit (2,399), Sami (1,991), Eskimo (1,750), Udeghe (1,657), Tubalar (1,565), Yukagir (1,509), Ket (1,494) and Chuvan (1,087);

- below 1,000 – Chelkan (855), Tofalar (837), Nganasan (834), Oroch (686), Chulym (656), Aleut (540), Kamchadal (2,293), Negidal (567), Orok /Ulta/ (346), Taz (276), Entsi (237) and Kerek (4).

Though Russian authorities of all levels pay special attention to the languages and cultures of those entities, the risk of their extinguishing should not be underestimated.

People with a vague idea of Russia's multiethnicity may think that minority languages are endangered because ethnic Russians have been assimilating their speakers for several centuries. This is not quite so for the Far Northern indigenous ethnic minorities, who are mostly assimilated by larger minorities. The Kerek, Koryak and Chukchi languages, all of the Chukchi-Kamchatka group of Paleo-Asian languages, make a good example.

The Kerek, a Paleo-Asian ethnic entity, live in the Chukchi Peninsula in Russia's Far Northeast. Only four said they were Kerek during the 2010 national population census. There were eight in 2002, compared to 102 in 1897

and roughly 100 in 1959. Archeologists date the profoundly original Old Kerek culture to the 1st half of the first millennium B.C.

Kereks lived in the 20th century in several villages side by side with the Chukchi, the largest indigenous ethnic entity in the peninsula which owes them its name. This tribe emerged at the turn of the 3rd millennium B.C. The Chukchi are only a small ethnic minority on the scale of entire Russia while they are a huge, mighty tribe according to local standards. Their number has been increasing lately: the 2010 census reported 15,908 as against 15,767 in the 2002 census.

Naturally, the Chukchi assimilated Kereks though the latter did not practice intermarriages. Kereks spoke basically Chukchi, using Russian to a smaller extent, while their native Kerek survived solely as passive knowledge in the preceding decades.

The Kerek language is genetically linked to Koryak, spoken in the Koryak Autonomous Area, which borders on the Chukchi Autonomous Area. Certain scholars regard Kerek as a dialect of Koryak, and the Kerek people were often considered Koryaks in the preceding centuries.

Koryaks have the same status as Chukchi – an indigenous Far Northern ethnic minority. There are 9,000 Koryaks presently. They live in high-density settlements in the north of the Kamchatka Peninsula, and speak Russian, for the most part. The Koryak language boasts only 2,000 speakers. It has no written variant due to their scarcity, and the language was first described as late as 1954-1956.

Unlike it, alphabets were elaborated for the Koryak and Chukchi languages in 1931.

Languages survive not only when spoken but also when studied. It is the best option to have an endangered language not only as an academic discipline but also as the tuition language. Understandably, it is impossible to teach all or at least several subjects in Koryak or Chukchi. However, they are studied, and so receive a new lease of life.

Koryak (to be precise, only one of its dialects) is studied in the 1st and 2nd years of primary school. Its teachers get education at the teacher training school in Palana, the administrative centre of the Koryak Autonomous Area. A total of 35 teaching aids have been published in Koryak. The presence of a great many dialects inhibits the development of Koryak as a literary language.

Chukchi, as the language of a larger entity, is a far more ambitious educational project. It was taught throughout the four years of primary school before a

resolution was endorsed in 1993 to teach Chukchi all the 11 years of ethnic secondary school. Study books have been made for 1st into 6th forms by now.

Chukchi is an academic discipline in the Chukchi Peninsula, particularly at the higher pedagogical college in Anadyr, its administrative centre, and in two other constituent entities of Russia – at the ethnic college in Chersky, Republic of Sakha (Yakutia) and the International Pedagogical University in Magadan.

Newspapers published in the Chukchi and Koryak autonomous areas have pages and supplements in Chukchi and Koryak. There are television and radio broadcasts in both languages, and original and translated books are published – both fiction and books on politics and social sciences. Literature in Chukchi and Koryak emerged in the 1930s and reached its peak in the 1970s. All educated Soviet people knew Yuri Rytkheu, a prolific Chukchi writer whose books were translated into Russian and from it into several European languages.

The Chukchi and Koryak languages are studied as full-fledged academic disciplines at one of St. Petersburg's most prestigious universities – the Far North Institute, a branch of the Herzen Russian State Pedagogical University.

Migrants' Languages

Mass migrations of the recent years account for the mounting presence in Russia of languages spoken in the former USSR – Azerbaijani, Armenian, Tajik, Kyrgyz, Uzbek and Moldovan, alongside Chinese and Vietnamese. Hundreds of thousands or even millions speak each of those languages in present-day Russia.

This presentation does not regard them as ethnic languages of Russia because they base on statehood outside Russia, and other countries are responsible for their survival: Armenia for Armenian, Azerbaijan for Azerbaijani, etc.

Conclusion

All I have said does not mean that Russia has no problems with the preservation, study, teaching and dissemination of languages. On the contrary, there are many obsessive problems – political, cultural, academic, educational, ethical and, last but not least, economic, considering the costs of language preservation.

These are two-fold problems: they concern minority languages according to the Russian national standard, which are often majority languages in the area they are spoken, particularly, in an autonomous constituent entity.

There are also problems with the preservation and study of the Russian language.

One of these problems is that certain ethnic autonomies promote their languages at the expense of Russian, on whose use and tuition limits are imposed. Local authorities' dedication to their language occasionally leads to absurdities, for instance, teaching it to ethnic Russian children since the age of four, when they don't properly speak even their native Russian.

We Russian nationals discuss these problems openly and widely, and they eventually find solution, though not so soon as we would like.

There is another, formidable problem that defies solution. The whole world shares it with us Russians. That is linguistic degradation, which grows worse with each generation of students. University professors complain: "Young people had difficulties with written essays ten years ago. Now, they cannot formulate an idea explicitly even in the oral form."

This problem does not concern only Russia or any other country – it concerns the entire world civilization.

PLENARY SESSION

Vitaliy KOSTOMAROV

President, Pushkin State Russian Language Institute
(Moscow, Russian Federation)

The Russian Language Brings People Together from the Atlantic to the Pacific

I was lucky to meet Rasul Gamzatov, a poet of genius. The memory of our heart-to-heart talks makes me bold to cite an impassioned line from his verse about Avar, his native language: "If my mother tongue vanishes tomorrow, I want to die today." He loved his native tongue, in which mother sang him lullabies and father told beautiful tales, and, with equal inspiration, he wrote about Russian, the language that told him about vast lands and made him treat all his fellow countrymen as friends:

"I walked across the mountains with Russian in my heart. It was a powerful language. <>Son of a mountaineer, I adopted Russian with my soul as my mother tongue. <>From the Baltic to Sakhalin, we share hearth and home as the offspring of one family."

Russian opened the world to the poet and gave him world renown.

It takes true Oriental wisdom to picture so graphically the sophisticated correlation between one of the world languages, whose command is a must in the globalization era, and native languages, of which there are close on 6,000 in the world (according to approximate statistics, which make no difference between language and dialect). Last month alone saw two landmark events dedicated to this theme: a UNESCO conference in Paris on languages and cultures in the contemporary world, and an online conference in Perm on the Russian language in the cultural dialogue. The present conference will certainly make an honourable contribution to the cause that will remain topical in the world for years ahead.

Any one language cannot yet aspire to the status of the only common language in the world. More than a billion people speak Chinese. Hindustani (Hindi plus Urdu) boasts a similar number of users. English comes next, with a far smaller number of native speakers. Despite that, it is studied everywhere and has conquered the global transport, commercial and IT spheres. Russian occupies an honourable place with 350 million users, and is taught in 67 countries. German, French, Spanish, Arabic and some other languages also

have an established place among the world languages, and one or more of them belong to mandatory school disciplines all over the world.

It would appear that it's more convenient to choose the most widespread language for an instrument of transnational communication – the language with the greatest number of students, which is determined by the economic, industrial, social and cultural situation. It is hard to underestimate the impact of religion, history, geographic neighbourhood, tradition, and the state of translation and book publishing on the role of particular languages. Different languages play the role of the common tongue in different parts of the world. The number of such languages is relatively small and historically changeable.

Russian is qualified as the national language of the multi-ethnic Russian Federation by its Constitution.

Interlinguistics, a notable academic discipline, regards the factors that impede the choice of a common language and focuses on the prospects for an artificial language whose introduction as an auxiliary common language would facilitate international communication while preserving all natural languages without granting special rights to any of them to prevent the privileged status of its native speakers. The acquisition of such an artificial language would be too good to be true in the present-day world despite certain success of Esperanto and other similar inventions.

High-falutin' eulogies of the common language are to its detriment. I made it a point to avoid them in my books *My Genius My Language* and *The Life of a Language: From the Vyatichi to the Muscovites*, and in many socio-linguistic articles. I emphasized there that all languages are equal to an unbiased linguist, and are equally beautiful as multi-faceted manifestations of human genius. At any rate, I did not think it was productive to compare the state structure of the multi-ethnic USSR to a patriarchal family headed by a father-like eldest brother, whose native Russian language was viewed as non-Russians' second mother tongue. Overly enthusiastic journalists used this doubtful metaphor instead of qualifying Russian as lingua franca. Worse still, the other ethnic entities and languages of the multi-ethnic country were regarded as "younger brothers" on the basis of their number and other formal criteria. That was why they were unequal in the distribution of radio and television air time, the circulation of books and periodicals, and the length of ethnic language classes. The concept of ethnic cultures as different forms of the same "socialist content" also clashed with the principles of democracy and equality.

The Russian language was a true instrument of communication and cooperation only when it was adopted consciously and, most importantly, voluntarily, and when its mandatory use did not threaten to oust other languages. A common

language is acquired and psychologically alien. That's what differs it from the native language, which is accepted literally with mother's milk. The mother feeds the baby's body to make it grow, and gets going the genetic programme that enables the baby to think and communicate. That is why native language is termed "mother tongue", not "father tongue". Even before the baby is born, lullabies acquaint it with phonetics, intonation and morphology – the basic material and technical elements of a language.

Importantly, these elements are only seldom obliterated. Usually they are manifest as foreign accent in other languages, however fluent the speaker might be in them. Vasily Abayev, an outstanding philologist of Osset ancestry, complained after he wrote several dozen books and delivered several hundred lectures in Russian, that he was afraid to calque his native tongue's constructions "stand a book on a shelf" and "lay a book on a table" and so said "place a book on a table/shelf". Ditmar Rosenthal, a peerless expert on the Russian language norm, who fluently spoke Polish and Italian, intoned his Russian, Polish and Italian speech as his native Yiddish. Bilingual poets Alexander Pushkin, Vladimir Nabokov and Isaac Brodsky wrote in Russian poems evidently superior to their English and French verse. Ivan Turgenev, with his brilliant French, said he could do creative writing only in Russian. There are certainly exceptions. Chinghiz Aitmatov wrote with equal perfection in Kyrgyz, Kazakh and Russian, and said it was pity that his English and German weren't fluent enough to try his hand at them.

To be sure, emotion and logic matter more than language in creative writing. Like Pushkin, every author should believe that "Long shall I a man dear to the people be for how my kindling lyre bid kindly feelings grow" and that it will bring him reward: "Tidings of me shall spread through all the realm of Rus and every tribe in Her shall name me as they speak" (translated by A. Z. Foreman).

I can only say that, putting it figuratively, a mother is irreplaceable though a stepmother might be more selfless and lovable. I mean that the native language remains an eternal substratum even when its speaker shifts entirely to another language. I will not dwell here on matters of tremendous interest: (1) Can one have two native languages? (2) Can an ethnic entity be bilingual? and (3) Are bilingual people bicultural at the same time?

Just as native and acquired languages, the common language accepts and disseminates everything good that appears in its users' native tongues in lasting interethnic contacts. This exchange enriches the common language as it promotes common features in other languages to make the basis of what linguistics term "language union" (jezični savez, Sprachbund).

The multilingual Encyclopaedia of Linguistic Terms (Simeon Rikard. *Enciklopedijski rječnik lingvističkih naziva*, tom I, Zagreb, 1969. s. 611) traces this term to the Prague linguistic school and defines it as the emergence of common features in the "Balkan union" of the Romanian, Bulgarian, Macedonian, Greek and Albanian languages.

As we adopt the idea and the term, we should notice that language unions are extremely diverse, and each of them is unique. The International Organization of the Francophonie unites 75 countries on several continents with total population exceeding 890 million, including 220 million French speakers, and highlights their cultural and linguistic diversity and the cementing influence of the French language (see its Secretary General's contribution in the collection *NET.LANG: Towards the Multilingual Cyberspace*, C&F éDITIONS, 2012).

The Eurasian language union spreads over a vast space from the Atlantic to the Pacific and from southern mountains to Arctic ice, where hundreds of ethnic entities have been coexisting and interacting since times immemorial. The Russian language, which Pushkin described as "imitative and liveable-with", has been its basis since the 15th–17th centuries. The diversity of its member languages and cultures, which belong to different systems, and their unequal developmental levels are salient features of that union.

Oleg Kuvayev's novel *The Territory* presents an exotic and controversial picture of one of the many areas in the Eurasian language union, with fur boots and coats, frosts and blizzards, dog sleighs, heroic acts, and fabulous riches scattered in a vast area. Another graphic example is Oldiria, a land that allegedly vanished like Atlantis. Many authors write about it.

Words are the simplest and the most spectacular testimony to the common features of interacting languages. Russian opened the culture of the Antiquity and the world of West European learning for all other languages in the Eurasian language union, while enriching itself with the names of ethnic dishes, dwellings, clothes and customs to spread them worldwide. Of even greater importance are semantic-cognitive and mental-linguistic structures, and the techniques of text production. It would be apt here to mention Nietzsche's intriguing hypothesis on "Maxim Gorky's two souls", which settled, in a way, the age-long disputes between Russian Slavophils and Westernizers.

In her doctoral thesis on "Intercultural Metaphors in Russian Creative Writing" (Moscow, 2003), Marina Subbotina tracks the contemporary general syntactic standards of narration down to Turan originals, i.e., contacts with Ugric-Finnish, Samoyed, Turkic and Mongol-Manchurian peoples since the 7th century – hence such stylistic devices alien to European languages but firmly

rooted in Russian as the fluid melody of narration, oblique and figurative authorization, or double negation.

Regrettably, this thematic range attracts only few researchers. The specifics and common features of the Eurasian union peoples are little-studied largely due to the assumption of their backwardness as compared to the American and West European linguistic and cultural standards. That is what makes so interesting and topical Ludmila Zamorshchikova's information testifying to the sophistication of Yakut, Yukagir and Evenk linguistic mentality (L. S. Zamorshchikova. Linguistic Consciousness of the Northern Peoples: Psycholinguistic Issues. *Language, Communication, and Culture*, 2012, No 1).

We can say assuredly that the extensive appearance in cyberspace of the latest facts testifying to linguistic and cultural diversity, particularly reflecting the associative networks of Eurasian material life, culture, philosophy, religion, customs and traditions, enriches our idea of global culture, united in its diversity, and the desired and actual patterns of global linguistic development.

Joseph MARIANI

Research Director, National Centre for Scientific Research &
Institute for Multilingual and Multimedia Information (IMMI)
(Paris, France)

How Language Technologies Can Facilitate Multilingualism

Abstract

The issues of multilingualism are many and the need for multilingualism is important, in Europe and internationally, both for preserving cultures and languages and for enabling communication among humans speaking different languages. The cost of multilingualism is however enormous, and cannot be covered by human forces only. Language Technologies can help forming a response, but are still under development despite their increasing penetration in daily applications. Furthermore, they are only available with a sufficient quality for a small fraction of the languages spoken worldwide. Increasing their quality and enlarging the language coverage would require infrastructure development, production of the resources needed to conduct research for those different languages and evaluation of the resulting language processing systems. Some major companies, mostly US, now recognize the importance of multilingualism for conducting their commercial business and invest a lot in that area, but they mostly address languages of economic interest. Some national and EU community programmes also support this domain, but suffer from a lack of scale, continuity and cohesion. This effort deserves to be coordinated among nations and international organizations, such as UNESCO, in order to facilitate multilingualism in Europe and globally, and avoid enlarging the digital divide.

1. Introduction

Since the divine punishment of Babel, mankind must live with the wealth of a multitude of languages and cultures. The difficulty and costs of sharing information and communicating, despite the language barriers, while preserving these languages, could benefit from the support of automatic language processing systems (that we will call Language Technologies), which are the object of a large research effort, although still insufficient and insufficiently coordinated.

2. The Issues of Multilingualism

The issues of multilingualism are twofold.

First, to take care of preserving cultures and languages, i.e. to allow citizens to express themselves in their first language. This question takes on a particular depth in the context of the construction of Europe, given the strong linguistic diversity within a single political entity. A study conducted for the European Commission (EC) shows that 90% of European citizens questioned prefer to find websites in their native language rather than in a foreign language. One can also note that it is currently estimated that less than 30% of the Web is in English, a proportion that has declined sharply from a rough estimate of 50% in 2000. 50% of European citizens speak only one language and when they speak a second one, it is not necessarily English. Only 3% of Japanese speak a foreign language. In India, less than 5% of people fluently speak English. Preserving languages and, through them, their corresponding cultures responds to a strong demand from citizens.

The second challenge is to enable communication among humans, usually in the framework of common democratic structures. We are facing it in the European Union (EU), where, with the recent expansion, there are now 28 Member States and 24 official languages, representing 552 language pairs. If one considers all the European languages, one can count more than 100, which represents more than 10,000 pairs of languages to translate! The European Commission employs more than 2,500 translators who translate about two million pages per year. This covers only a fraction of the needs. To cover the totality would require 8,500 translators to process 6.8 million pages annually. Note that EU linguistic diversity represents 30% of the budget of the European Parliament, or about 300 million euros per year, with the use of 500 translators and interpreters. The estimated total cost of multilingualism for the European Union is a little over one billion euros per year; but considering the number of Europeans, that represents only 2.2 euros per citizen per year, which ultimately is not prohibitive. The same study conducted for the EC showed that on the economic side only 30% of EU citizens would accept to buy goods over the Internet in a foreign language, while, on the cultural side, 80% of those citizens think that web sites existing in their language should be translated to foreign languages. A similar situation exists within some multilingual nations, like India, but also internationally, with about 6,500 languages that are spoken and more than 40 million pairs of languages to translate... And a simple statistic: at present YouTube, every minute, uploads one hundred hours of new videos in all languages.

3. Needs Related to Multilingualism

At the European level, the needs related to multilingualism are very numerous: the European Digital Library (Europeana), which included, in 2013, 23 million documents in 26 languages and for which it is necessary to provide multilingual and cross-lingual tools to enable access to information for all, whatever the language in which the information has been encoded. The European Security Agency (ENISA) plans to produce a multilingual platform for alert and information exchange for the EU Member States. The European Patent Office has reduced, according to the London Protocol, the number of working languages to three (English, German and French) for reasons of cost, while they receive 265,000 patents per year in their 28 official languages plus Russian, Chinese, Japanese and Korean. It is estimated that translating their complete 10 million patents portfolio in those 32 languages would require 1,500 years for a team of 1,000 translators! For the same reasons of cost and feasibility, English tends increasingly to become the only working language in meetings of the European Commission, of the European Parliament or of the European Court of Justice. In 1997, 45% of the EC source documents to be translated at the EC were written in English and 40% in French, while in 2007, 72% of those documents were in English and only 12% in French!

Such needs respond to a real democratic necessity, to be met more generally at the international level. If we take the example of Internet governance within the UN Internet Governance Forum (IGF), only English is accepted as a working language, and a lively debate concerned the possibility of using different spellings and different accents in the domain names. The World Digital Library in UNESCO reached 10,000 documents from 80 countries in February 2014[2]. Dubbing and subtitling of audiovisual works; writing technical manuals, in the aerospace or automotive industries, or instruction manuals for consumers; conducting commercial business at the international scale; live super-titling of works of performing art; translation of texts, videos and radio or television programmes that are innumerable, and in all languages; simultaneous interpretation in military or sanitary operations which take place around the world (such as the ones following the Haiti earthquake) and at multiple meetings, conferences or workshops; interpretation of courses, with the coming of MOOC (Massive Open Online Courses)... Think also of the urgent needs related to scientific articles written in a mother tongue, which are diminished markedly due to the overvaluation of English by bibliometrics, risking the loss of specialised terminology in other languages. In 1980, 85% of

---

[2] *http://www.wdl.org/fr/.*

the publications referred in the Science Citation Index (SCI)[3] were in English, 4% in French and 4% in German. In 2000, 97% of the publications were in English, and only 1% in French or German [Bordons and Gomez 2004].

Add to this picture the many needs related to the accessibility of information by the visually or hearing impaired, requiring the translation of information from one medium to another (written to oral, oral to written, oral to gesture (sign language)), and more generally to the accessibility of information by people who do not speak fluently the language in which it was encoded, including, notably, migrants.

4. Findings

The extent of these needs shows very well that they cannot all be covered by existing or even future human resources of professions dealing with language processing.

We should understand that multilingualism is not a top priority in any economic sector. If we ask the boss of a big company what is his/her priority, none will say it is multilingualism. But if we add up the priorities in each area where it is necessary to take it into account, then we reach a very large sum. This therefore requires, in our opinion, thought and political action to bring out this awareness and provide appropriate responses.

Even when multilingualism is seen as a necessity, its cost is too important. It is this gap that calls for the development of Language Technologies and their utilisation, but only when their performance is up to the needs of target applications.

If a language does not benefit from the availability of Language Technologies, it won't be used in voice operated high tech devices (such as car GPS, Smartphone interaction, Internet search engines, or Emergency Calls) where it will be replaced by another language and may thus get in danger of "digital extinction". Meanwhile if it benefits from Language Technologies such as Machine Translation, it will keep being used even in confrontation with much more widely used languages.

It should be noted that currently, Language Technologies have not yet reached maturity for all languages, with strong imbalances among languages, and that they cannot replace humans. For example, automated translation is not good enough to translate literary works or, in general, texts which require high quality translation. This must be said clearly. But on the other hand, it can

help a human translator in his or her work and has a sufficient quality to give an approximate translation, of web pages for example, thus meeting the needs of the general public.

Language Technologies can more fully participate in solving the issue of multilingualism, which justifies drawing attention to their merits, especially in the funding of large research programmes.

5. Language Technologies

Language Technologies are said to be *monolingual* when they handle a single language, *multilingual* when the same technology processes several (individual) languages, or *cross-lingual* when they allow for switching and transferring from one language to another.

Language Technologies cover the processing of written language, whether monolingual (morphosyntactic and syntactic analysis; text understanding; text generation; automatic summarization; terminology extraction; information retrieval; systems that respond to questions (such as IBM Watson[4]), etc.) or cross-lingual (automatic or computer-aided translation; cross-lingual information retrieval, etc.).

For the processing of spoken language, there are also monolingual technologies (speech recognition and understanding; speech-to-text transcription (textual transcription of what has been said); speech synthesis; spoken dialogue; speaker recognition, etc.) and cross-lingual ones (identification of a spoken language, speech translation, real-time interpretation, etc.).

Finally, it includes the processing of sign languages (recognition, synthesis and translation).

These technologies can be intermedia, i.e. translating from one medium to another, with numerous applications to enable accessibility for the disabled (Text-To-Speech synthesis for the visually impaired, automatic transcription (subtitles or supertitles), aids to lip reading, Sign Language processing for the hearing impaired, voice commands for the motor-impaired).

Numerous resulting applications are now in everyday use, such as, regarding written language processing, spelling and grammar checkers, monolingual and cross-lingual search engines, online machine translation, etc., and, regarding spoken language processing, talking GPS systems, voice dictation, automatic transcription and indexing of audio-visual content, spoken translation, etc. This list shows that many of these existing applications are related to linking

[4] *http://www.ibm.com/smarterplanet/us/en/ibmwatson/.*

spoken and written language (transcription of speech into text, speech synthesis from text). Spoken dialogue systems, including voice recognition and synthesis, are also growing, but still in limited applications: spoken interaction on Smartphones (such as Apple SIRI), Call centres, tourist or public transportation information, etc.

6. Language Resources and Evaluation

It is crucial for conducting research aiming at developing Language Technologies to provide a base including both language resources and evaluation methods for the technologies that are developed.

With regard to language resources, the data (corpora, lexicons, dictionaries, terminology databases, etc.) are both necessary for conducting research investigations in linguistics and for training automatic language processing systems that are based in most cases on statistical methods. The greater the amount of data, the better the statistical model and therefore the better the system performances. The interoperability of language resources also invites us to think more deeply on the standards to be put in place in order to organize, browse, and transmit data.

It is also necessary to have a means for evaluating these technologies in order to compare the performance of systems, using a common protocol with common test data, in the context of evaluation campaigns. This allows for comparing different approaches and having an indicator of the quality of the research, of the advances of technology and of its readiness compared with the needs of targeted applications. We now speak of "coopetition" – a mix of international competition and cooperation – and this has become a way to carry on technological research. The Defense Advanced Research Projects Agency (DARPA) of the Department of Defense in the United States was the initiator of this approach in the mid 80s, through the National Institute of Standards and Technology (NIST) [Mariani 1995][5].

A similar approach was used to monitor progress in machine translation (MT), using the BLEU metrics, proposed in 2000 [Papineni K. et al. 2001], whereas the research had been conducted in MT for about fifty years without systematically measuring the quality of results to guide research. This measure is based on a rudimentary comparison between the results of the systems and the translations by human translators.

[5] *http://www.cslu.ogi.edu/HLTsurvey/.*

Referring to the initial issues, we can pick up the two key elements that are necessary for a Language Technology policy: the availability of monolingual resources and technologies in each language, in order to ensure the preservation of culture (and therefore of languages) and, at the same time, the availability of cross-lingual resources (such as parallel corpora) and technologies for each pair of languages to be processed, in order to enable communication between humans.

7. The Digital Divide and Language Coverage

There is currently a two-speed situation and a "digital divide" between languages for which technologies exist, and others. This is related to the "weight of languages"[6]. It should be noted that 95% of languages are spoken by only 6% of the world population and may not represent an economic interest for companies. Some linguists believe that 90% of languages will have disappeared within a century. We can therefore classify languages according to the data and automatic processing systems that exist for these languages: whether they are well, less or not at all "resourced, or indeed if they have only an oral tradition and no writing system at all. Only 1-2% of the ca. 6,500 languages spoken worldwide benefit from Language Technologies. The availability of data is crucial for the development of usable systems, often based on statistical approaches. Machine translation therefore requires parallel corpora, whose number is reduced. Therefore we try to overcome this gap by developing methods using noisy parallel corpora, comparable corpora (texts dealing with the same topic in different languages) or quasi-comparable corpora, which are more readily available, thanks especially to the extension of the Web.

In order to resolve this digital divide, how can we take into account "minority" languages, regional languages, languages spoken by migrants, foreign or regional accents? Who bears the cost when these languages are of no economic or political interest, or are unrelated to armed conflicts or natural disasters that justify addressing them? How to ensure that citizens in a community of states are able to communicate among themselves? How to reduce the risk of conflicts and crises by allowing exchanges between people? This is now a major social and political issue, which is the subject of much debate.

Thus, the International Forum of Bamako, organised in January 2009 in pursuit of the outcomes of the World Summits for the Information Society in Geneva (2003) and Tunis (2005), concluded on a commitment to promote ethical use of

---

[6] *See [Gasquet-Cyrus, Petitjean 2009].*

information in its linguistic dimension, allowing mother tongue education and ensuring the existence of a multilingual cyberspace, both in terms of content availability on the Web and of technologies to access it.

8. Research Efforts in the Domain

To produce the language resources and technologies that are needed to address multilingualism, different initiatives can be identified:

- those of large companies such as Google, IBM, Facebook, Apple, Amazon, e-Bay or Microsoft;

- national programmes in some countries, with different objectives: to process internal multilingualism (TDIL in India, NHN in South Africa); to understand foreign languages for geopolitical reasons (GALE or EARS in the United States, funded by the Department of Defense (DARPA)); to ensure the use and promotion of a national or transnational language (TechnoLangue for French, STEVIN for Dutch/Flemish); or to maintain a place in an economic and cultural competition (Quaero in France);

- R&D programmes of the European Commission;

- international efforts to network the actors of the field, to better coordinate activities and promote greater sharing of resources ((Oriental) Cocosda, CLARIN, FLaReNet, META-NET, etc.) and the establishment of distribution agencies for linguistic resources, such as LDC in the United States or ELRA in Europe.

8.1. Producers of Information Technology

It must be underlined that large U.S. companies in the information technology sector make a major effort in multilingualism and cross-lingualism. Thus, the Google search engines work in 145 languages (national and regional), and Google has made available "free" tools for machine translation and cross-lingual information retrieval online: in June 2014, 80 languages (including Catalan and Galician) and 6,320 language pairs were available on the Internet, and on smartphones (including more than 17 languages with voice input and 26 languages with voice output, and several varieties of English, Spanish and Chinese), and Google targets 100 languages, i.e. 10,000 language pairs, by 2015. Google Translate has 200 million users and translates the equivalent of 1 million books per day, that is more than what professional translators do in a year. It provided "for free" the automatic translation of patents in 32 languages for the European Patent Office, after additional training on their corpus. As of

April 2013, Google Books contained 30 million documents in 46 languages. In December 2010 Google provided statistics on the evolution of human language from a corpus of 500 billion words (including 361 billion words in English and 45 billion words in French and Spanish). Also Microsoft provides the MS Word spelling checker in 126 languages (233 if we consider regional variants) and a grammar checker in 6 languages (61 if we consider regional variants). The Apple Siri Speech Interface on iPhone is available in 8 languages, and 19 language varieties (Chinese (3), English (4), French (3), German (2), Japanese, Korean, Italian (2), Spanish (3)). And Facebook, Amazon, IBM or e-Bay invest a lot in that area.

8.2. National Programmes Addressing the Issue of Language Technologies to Help Multilingualism within a Country: TDIL in India, NHN in South Africa

Major programmes were launched as part of public policy. The TDIL[7] programme (Technology Development for Indian Languages) is an important programme, which is one of ten priorities of the Indian national programme on the information society. The target is to process (Indian) English and the 22 "constitutionally recognized" Indian languages (Assamese, Bangla, Bodo, Dogri, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Malayalam, Manipuri, Marathi, Nepali, Odia, Punjabi, Sanskrit, Santali, Sindhi, Tamil, Telugu, Urdu), with several Language Technologies: machine translation, Text-To-Speech synthesis, speech recognition, search engines, optical character recognition (OCR), spelling checkers, language resource production; all this for the group of 23 languages. A comparable programme (NHN[8]: National Human Language Network) is taking place in South Africa for the automatic processing of the eleven national languages (Afrikaans, (South African) English, isiNdebele, isiXhosa, isiZulu, Sepedi, Sesotho, Setswana, SiSwati, Tshivenda, Xitsonga).

8.3. Actions of the European Union

From 2007 to 2010 the European Union benefited from having a commissioner specifically for multilingualism[9], who established a High Level Group on Multilingualism, which produced a report that was presented to the EU Parliament and the European Council in September 2008. Within its presidency of the European Union, France organised in September 2008 the *Etats-Généraux du Multilinguisme* (Multilingualism Summit) at La Sorbonne

---

[7] *http://tdil.mit.gov.in/.*

[8] *http://www.meraka.org.za/nhn.*

[9] *http://ec.europa.eu/commission_barroso/orban/index_fr.htm.*

(Paris) that was followed in November 2008 by a resolution of the European Council of Ministers on multilingualism, taken up by the European Parliament in March 2009.[10] The idea of a "Single European Information Space" was highlighted. More recently, France organized the follow-on *Etats Généraux du Multilinguisme dans les Outre-Mer* (Multilingualism in Overseas Summit*)*, addressing the languages spoken in French overseas [DGLF2 2011].

The European Commission has supported several important projects on multilingual technologies under the 6th Framework Programme for Research and Development (CLEF, TC-Star, CHIL, etc.). In particular, the TC-Star[11] Integrated Project covered speech translation in three languages: English, Spanish and Chinese, through an application performing automatic translation of the speeches at the European Parliament. Working in this context is very interesting because all the necessary resources exist at the European Parliament: members' speeches in their own language, their (speech) interpretation in different languages of the Parliament, their transcription into written form, and the translation of the transcripts in different official languages. Thus, these data allow for training the automatic interpretation systems, including recognition in the source language, translation from the source language to the target language, and speech synthesis in the target language, thus utilising both monolingual and cross-lingual technologies. TC-Star has also produced and distributed a report on the status of Language Technology in Europe [Lazzari, Steinbiss 2006].

In the seventh European Framework Programme, FP7 (2007–2013), this area was mainly conducted by the "Language Technology, Machine Translation" Unit. In addition to R&D projects, an infrastructure and two networks have been established: CLARIN (Common Language Resources and Technology Infrastructure)[12], FLaReNet (Fostering Language Resources Network)[13], and META-NET (Multilingual Europe Technology Alliance)[14].

CLARIN is an infrastructure supported by the programme ESFRI (European Strategy Forum on Research Infrastructure) of the European Commission. Its objective is the distribution of language resources and tools for the Human and Social Sciences.

---

[10] *http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P6-TA-2009-0162+0 +DOC+XML+V0//FR&language=FR.*

[11] *http://tcstar.org/.*

[12] *http://www.clarin.eu/.*

[13] *http://www.flarenet.eu/.*

[14] *http://www.meta-net.eu/.*

FLaReNet is a Thematic Network supported under the e-Content European Programme, with a budget of €0.9 million over 3 years (2008–2011). Its purpose was to serve as a think tank for the promotion of language resources in European programmes.

The META-NET Network of Excellence was established within the T4ME (Technologies for a Multilingual Europe) project. This project had a budget of €6 million over a period of 3 years (2010–2013) and was structured in three parts: i) pushing the research frontiers in machine translation, ii) establishing an Open Resources Infrastructure (META-SHARE), including the production, annotation, standardisation, validation and distribution of language resources, and the evaluation of Language Technologies, iii) conducting a reflection on the place of multilingual technologies in the context of the next EC Framework Programme. A series of White Papers has been produced covering 31 languages. Each volume describes the status of the language and the level of technologies addressing that language in four areas (Text analysis, Speech processing, Machine Translation and Language Resources). It showed that 21 of those languages are under-resourced, as it appears in Language Matrices and Language Tables providing a comparison across the languages[15], and are therefore in danger of digital extinction. META-NET also produced a Language Technology Strategic Research Agenda providing recommendations (including the use of technology evaluation and the necessity of sharing the research effort with Member States through the existing EC organizational instruments) and the corresponding roadmaps in three areas (Translingual Cloud, Social Intelligence and Interactive Assistants) for the Horizon 2020 EC Framework Programme (2014–2020).

9. European and International Perspective

The resolutions of the European authorities demand a major effort to process all European languages, national and regional. However, if one considers the number of languages or language pairs that are to be addressed, and multiply it by the number of technologies, we see that the size of the effort is probably too large for the European Commission alone. It would therefore be interesting to share this effort among Member States, or regions, and the European Commission, in perfect harmony with the "principle of subsidiarity".

Language Technologies are well suited for a joint effort. The European Commission would have the primary responsibility for overseeing and ensuring coordination of the programme (management, provision of standards,

---

[15] *http://www.meta-net.eu/whitepapers/key-results-and-cross-language-comparison.*

technology evaluation, communication...) and of developing core technologies around language processing. Each Member State would have as a priority to ensure the coverage of its language(s): to produce the language resources essential for the development of systems (corpora, lexicons, dictionaries), and to develop or adapt technologies to the specificities of its language(s). This model would be easily adaptable to an international effort, combining the efforts of the participating countries and of international organisations.

Unfortunately, until now the topic of Language Technologies has been regrettably considered just as one research area among many others in Europe, not as an essential element of European construction, requiring a high priority effort to handle the corresponding issues. This weakness is all the more dangerous given the liveliness of the Union and its needs to increase economic, informational and cultural exchanges between countries, and to address the citizens of each Member State and help them in their communication.

Despite the recommendations contained in the META-NET White Papers and Strategic Research Agenda, the new H2020 programme doesn't respond to those needs. The size of the effort on Language Technologies is still insufficient with a budget of about 30 million euros for 2014–2015. Written and Spoken Language Processing are now addressed in two different Units: "Data Value Chain" for the former, and "Creativity" for the latter. In the first 2014 Call for Proposals, only Machine Translation is addressed for a budget of 15 million euros, and English, French and Spanish are not eligible either as source or as target languages as they are considered as sufficiently covered! Research on Spoken Language processing goes with multimodal and natural computer interaction with a budget of 7.5 million euros. The political dimension of Language Technologies for Europe is not yet recognized, apart from the inclusion of a "Translation Cloud" as a Digital Service in the Connecting Europe Facility (CEF) programme. But the budget attached to this Public Procurement action is only 4 million euros for the two first years (2014–2015), and spoken translation is not considered, as being too "immature"!

Let's hope that the political awareness of the issues attached to multilingualism will see one day Language Technologies receive adequate attention as a major issue at the European and international levels.

10. Conclusions

Language Technologies are the only way to allow for full multilingualism in Europe and worldwide. They are presently available for a small set of languages and the other languages are in danger of digital extinction. It is therefore proposed to coordinate the efforts of States, even regions, and international

organisations, involving industry and public research laboratories. Care should be taken to produce for each language the language resources needed, and organise the research effort in an open way, based on the interoperability and objective benchmarking of technologies. UNESCO could assume a major role in the general coordination of those efforts, and ensure that no language is left behind.

We could then add a nod to the famous phrase of Umberto Eco, by saying: "Translation is the language of Europe… with the support of technology", and extend this assumption to the global village.

References

1. Bordons, M., Gomez, I. (2004), Towards a single language in Science? A Spanish view, Serials, 17(2), July 2004.

2. Cencioni, R., Rossi, K. (2008). Language based Interaction, EC-ICT Conference, Lyon, 26 November 2008.

3. DGLF2 (2011). Etats Généraux du multilinguisme dans les Outre-mer à Cayenne, Guyane, ISSN 1955-2890.

4. Eco, U. (1993). *La langue de l'Europe, c'est la traduction*, Assises de la traduction littéraire, Arles, 1993.

5. Gasquet-Cyrus, M., Petitjean, C. (eds.) (2009). *Le poids des langues*, L'Harmattan, 2009.

6. Koehn, Ph., Birch, A., Steinberger, R. (2009). 462 Machine Translation Systems for Europe, Machine Translation Summit XII, pp. 65–72.

7. Lazzari, G., Steinbiss, V. (2006). *Human Language Technologies for Europe*, TC-Star Report, April.

8. Mariani, J. (ed.) (1995). Evaluation chapter in: *Survey of the State of the Art in Human Language Technology*, R. A. Cole, J. Mariani, H. Uszkoreit, N. Varile, A. Zaenen, A. Zampolli, V. Zue (eds.), Cambridge University Press, 1995.

9. Papineni, K., Roukos, S., Ward, T., Zhu, W.-J. (2000). BLEU: A Method for Automatic Evaluation of Machine Translation. In: *Proceedings of the 40th Annual Meeting of ACL*, Philadelphia, PA.

Michael GIBSON
Senior Sociolinguistics Consultant, SIL International
(London, United Kingdom)

A Framework for Measuring the Presence of Minority Languages in Cyberspace

The Internet can be seen as a refuge for minority languages to find a place of self-expression, but also as a place of danger for the same languages, as speakers might be encouraged to switch to other languages, as they find that their own language does not serve them for the things they want to do. It seems in fact that both visions – of the Internet providing opportunities for language communities to use their language in new places and ways, but also being the means of faster introduction of a more powerful language to the detriment of the mother tongue – each contain more than a grain of truth. As our interest here is on steps towards creating safe niches for minority languages, we will focus on the positive aspects of Internet use, and how analysis may assist in identifying the optimal steps to strengthen a language's digital vitality. Our primary focus is also on lesser-spoken languages, rather than languages which count multiples of millions of speakers. Not all types of web presence are of the same nature, and this paper seeks to provide a tentative framework of these different sorts of web presence, along with reflection on the different impacts each type may have.

One of the primary advantages of the Internet for minority languages is the relative ease of content production, with a blog, for example, needing much less infrastructure than a book to produce. The Internet can also be an ideal medium for collaboration between speakers in different locations, thus enabling members of a diaspora, to which the more educated members of the community may belong, to play a full part in digital language activities. Much more detailed discussion of how the Internet can be helpful to minority languages can be found in [Vannini & Le Crosnier 2012].

Minority languages tend to exist in multilingual environments, where within the society different languages are used for different purposes. This is in opposition to monolingual environments that exist as the majority model in, for example, most European nations. Here one language, such as English or French, is often used for all possible functions within the society. But in parts of Kenya, for example, it would be common practice for family communication to take place in one language, such as Kikuyu, interaction on the street to be in the inter-ethnic language Swahili, and work-related correspondence to be

in English; such scenarios are the norm for members of minority communities. Multilingualism can be additive, where the learning of a new language does not threaten the maintenance of the mother tongue, or subtractive, where the learning of a new language results in losses of competence in the community language. As such, like the Internet, multilingual practices are not necessarily dangerous to minority languages, but in many cases they are also the intermediate stage through which language loss and shift occur. In and of itself, the greater ease that the Internet introduces for contact with other languages does not endanger minority languages, being of the additive type of multilingualism. But it would be naïve to suggest that this process is not also sometimes part of the process of language shift.

Language Vitality Frameworks

There are various measures of overall (rather than digital) language vitality, starting with Fishman's [1991] *Graded Intergenerational Disruption Scale (GIDS)*. This has since been further developed into the *Extended GIDS (EGIDS)* by Lewis and Simons [2010]. In these two scales, the central question relating to vitality is the age of the youngest generation of speakers, which entails the question of whether the language is being transmitted to children in the home, in which case the language can be regarded as vital. It would be fair to say the focus is on current language practices, and that any prediction is extrapolated from this current practice. The other dominant framework is that of UNESCO [Brenzinger et al. 2003], which takes into account a broader range of factors into the basic calculation, some of which are social rather than linguistic, and are seen as likely correlates of the language's future. The latter two frameworks are freely available on the Internet (see references). All these frameworks enable cross-comparison of different languages using the same criteria, enabling those concerned with the languages to understand the situation and take better-informed remedial steps (see also Lewis & Simon's [2011] *Sustainable Use Model,* which makes recommendations of the appropriate type of activity at each level of vitality).

Concerning digital vitality, the dominant model is found in Kornai's [2013] paper *Digital Language Death*, which aims to adapt the EGIDS in particular to presence and absence of languages on the Internet. Digital vitality generally depends on the vitality of the language in broader society; the latter is a necessary but not sufficient factor for digital presence. However, as intergenerational transmission is not the primary vector of digital presence, and digital presence can be more easily tracked through web crawling and automatic language identification, a significantly different scale, based on use, emerges.

His conclusions are not encouraging, for example he claims that the "vast majority of the language population, over 8,000 languages, are digitally still, that is, no longer capable of digital ascent" [2013: 1]. Kornai's primary interest is at the bottom end of the scale (the difference between dead and alive), but as his scale is composed of four levels, and is the basis of what we suggest in this paper, basic details of each level will be given here.

Kornai's Scale of Digital Presence

Thriving	**T**
Vital	**V**
Heritage	**H**
Still	**S**

Thriving is the top end of the scale, with large use by both native and foreign speakers, and extensive computer support from both Microsoft and Apple [ibid: 5]. **Vital** languages do not have such support, but are still "used for communication by native speakers" [ibid: 5]. Such communication by native speakers is lacking for the **Heritage** category, which covers cases where there are language materials, but these are "languages that are digitally archived" [ibid: 5], covering both currently vital languages where outside scholars have documented the language, and languages which are no longer spoken, and the "digital presence is *read only.*" As Kornai [ibid: 2] correctly comments "such efforts, laudable as they are, actually contribute very little to the digital vitality of endangered languages." More information on factors relevant to strengthening heritage status can be found in Gibson [2012a]. Such activity can be helpful for purposes of communal identity maintenance and connection with tradition; worthy activities, which however do not equate with digital vitality. Digital presence is only truly vital when there is writing by the community. The final category, **Still**, is where there is no observed use of the language, and, according to Kornai [ibid: 1], in such cases the language is "no longer capable of digital ascent."

Given that Internet usage is still increasing, and some parts of the world even now have little Internet access, we raise the question of whether this judgement of the digital stillness of the majority of the world's languages might

be premature in some cases. In particular, the rise of the smartphone and of the related activities of texting and social media private messaging are still ongoing in many parts of the world; they are also more difficult to observe by an outsider, as such use remains private. Other activities such as posting Facebook statuses and responding to them are public to varying degrees, and may be places where web crawling may yet show us signs of nascent digital vitality. It is our goal here to look at the likely routes of digital ascent, and to expand Kornai's framework to account for these intermediate stages. In doing so, some languages will nevertheless be judged to be incapable of digital ascent, and if we are able to make this judgement, it may help those of us who are concerned about the fate of the world's minority languages to concentrate our efforts on working with communities where digital ascent is still a realistic possibility.

Language and the Internet

Until the arrival of the Internet and the mobile phone, Abercrombie's [1963: 14] insightful comment that "writing is a device developed for recording prose, not conversation" held not just for its development but also its practice. Multilingual societies tend to reserve different sociolinguistic domains [Fasold 1984: 183] for different languages, and writing, being permanent and non-conversational, tends to trigger the use of more prestigious languages. Thus it is not normally a preferred domain for the vernacular, and a pattern where speakers of vital minority languages write in another language is not rare. This can be a challenge for those wishing to see greater use of vernaculars in writing.

Now with the new uses of writing that arrived with the mobile phone and Web 2.0, where much content is user-generated rather than published by brokers of the word such as newspapers and publishing houses, writing is no longer permanent (especially in some apps such as Snapchat, where the written message disappears soon after being sent), and is often conversational. This seems to account for the fact that textspeak (see [Crystal 2008]) is often deliberately non-standard, and, for example in countries as diverse as Tunisia and Kenya, will often be a place where speakers of non-standard dialects or minority languages are most likely to use them in writing. Coulmas [2013: 131] adds that "the fact that the telephone is the prototypical communication tool of oral-only exchange may have contributed to the hybrid character of instant messages ... by way of incorporating features of conversational performance into writing once the handset was equipped with a visual display." Similar patters can be seen in Facebook status updates, generally not motivated by

language activism, but because expressions of solidarity which go together with conversation increase the use of the non-standard and non-prestigious.

As such, in the case of minority languages, texting and messaging will be the areas where the psychological barriers to writing in the heart language are lessened, and we are most likely to see the beginnings of vernacular literacy. These new sociolinguistic domains, brought about by technological developments, have changed the nature of writing – it is no longer necessarily permanent or incompatible with spontaneous conversation. And here we can see a place where the impact of digital practices can extend beyond the digital sphere; texting in a mother tongue does not only encourage other digital literacy , but also provides a broader model of writing a minority language. As such, we argue that without texting or messaging, other forms of writing will fail to take root and the language will be incapable of digital ascent – if a language is not written in vernacular domains, which are its most natural homes, how will it be used in more formal ones?

Extending the Framework

However, under Kornai's framework, a language or variety which is being used for texting and messaging, but not on the open Internet, would still be categorised as still. This stage is what we call **emergent**. But we do recognise that if there is widespread use of a language on mobile phones, it would be unlikely to find none on the open Internet, even if this is not the primary place that it will be found. And here the question of perspective comes in. While working from above, looking at the macro picture, the use of some languages in cyberspace will be deemed as insignificant. While working from within the language community however, even such apparently minimal use may have significant impact on the literacy practices of that community, and that is the perspective I am wanting to foreground here – how can linguists (and others) work with communities to help them achieve their goals for written communication?

There are, however, some languages which show almost no sign of digital ascent. Here we mention some factors which play a role in whether digital use may start or not. Whether these factors are in place has a role in whether the language will be judged as **still** or **latent**.

- *Active intergenerational transmission.* As mentioned by Kornai, if the language is not being used as a medium of communication in the community, then digital practices will not progress towards the vital stage.

- *An available model of writing in the language*. Some sort of written use of the language often serves as a model for other uses. It is only an occasional activist who writes in a language that they have not seen written. Use of the language in education, whether as a medium of instruction or a subject of study can serve here, as can the presence of literature such as religious texts or worship aids, e.g. hymn books. This model of writing will not necessarily be followed precisely; variant spellings will be common, and will often reflect the speaker's own dialect, or the latest innovative youth usage. So, for example, seeing written Swahili or vernaculars in Kenya, where they are used in both religious worship and to a limited extent in education, seems to have encouraged widespread informal digital uses of these languages, and their variants such as the Swahili-based youth code Sheng [Githiora 2002]. Writing practices in a closely related language can also serve as a model for speakers to emulate, and in fact vernacular writing does not necessarily respect pre-ordained boundaries between or definitions of languages.

- *Sufficient software support to write the language easily*. Whereas we saw that digitally thriving languages have OS support, the level of support here is not equivalent. Where their own script has not been available, speakers of languages written with non-Roman scripts have shown themselves willing to write in Latin characters, for example in writing Hindi, Greek and dialectal/non-standard Arabic, where numbers have been used to represent sounds not handled well by the Latin script, in a style known as *Arabizi* [Randa et al. 2011]. As non-Roman scripts have become more widely available on a variety of devices, their use has unsurprisingly increased. But from this we can deduce that where the motivation to write in one's own language is high, speakers will find a way to minimise the challenges, happily departing from norms that do not suit them. So, in this case, sufficient script support may be present in a smartphone. Obviously, the better the support is for the language in question, the more this helps the written use of the language. The recent proliferation of smartphones and tablets, with touchscreen keyboards, makes localisation easier, as the technological backup to create different on-screen keyboards, such as those introduced by *Boite A Innovations* (http://www. boiteainnovations.com/index_en.php), is much less than that for creating a specialised physical keyboard.

The Proposed Framework

Thriving	**T**
Vital	**V**
Heritage	**H**
Emergent	**E**
Latent	**L**
Still	**S**

Under this proposal three of Kornai's categories are unaffected: thriving, vital and heritage. Our concern has been with the cases where there is little use or it is restricted to private domains of texting and messaging.

The **emergent** stage, which we have argued to be an essential and key stage for digital ascent to occur, is that of community use of texting and social media messaging. This tends to be driven by members of the community themselves, though language development projects may address issues of the writing system, dictionary and appropriate software support (for example K. David Harrison's online dictionary, including a keyboard, of Tuvan at http://tuvan. swarthmore.edu). We see these new domains of writing as an opportunity to further establish writing in the same languages. However the very advantage of these private conversational domains – their friendliness towards all that is vernacular – also represents a difficulty for those who wish to emphasise standardisation. This is typically a domain which does not submit to a standard, often being a place open to innovation and language mixing (which is a common feature of many youth-oriented codes such as Arabizi and Sheng, mentioned above). Those advocating for the use of minority languages often have a desire for a pure form of the language, with minimal influence from other languages, especially from those seen as a threat, such as English. We see this in the efforts to develop new vocabulary which may be at variance with community practice. Sometimes these interventions can be successful, but it is also possible that a strong emphasis on language purity can discourage use by younger speakers, who feel they no longer speak the language as it ought to be spoken. And a language not being used by young people has a perilous future.

In cases where a language has a lot of dialectal variations within it, or there are significantly different practices in urban contexts or among youth, emergent practice in vernacular writing can form the basis of new conventions (as in the decentralised conventions in Arabizi surrounding which number represents which sound). These practices in turn can be part of the development of recognition of different codes – such as in Nairobi, where many speakers differentiate between Swahili and Sheng, but not in fully conventionalised ways [Gibson 2012b]. There is also the possibility for finding common ground between what may have been viewed by some as different languages, such that an intermediate written form suits more than one community. At this point we need to note that which selection of varieties constitutes a language is by necessity a construct, primarily negotiated by speakers of those varieties, and so such definitions are sometimes fluid and dynamic. It is therefore possible that Internet usage will help define new varieties of language, even if the researchers are not committed to such varieties necessarily being defined as languages. But this does open up the possibility of a more democratised, less centralised way of defining language boundaries (if that is what we want to do), based on informal digital practices. The fact that these practices are unlikely to become fully standardised remains an issue to ponder further.

The **latent** stage is more difficult to justify empirically than the emergent stage, a point made by Kornai. From the point of view of data collection from web crawling, it would be an empty category. And yet from the community perspective, it represents a useful distinction between situations where digital ascent is possible (therefore at the latent stage) and where it is very unlikely (the still stage). For example, we may identify situations where there is no model for writing. Without that issue being addressed, the language will remain still. Furthermore, if the language is not being passed on to children in the home, any language activism or development activity will need to be focused on the transmission in the family. Without this, any digitally-based activities are doomed to failure, as there will be no community use behind it. Note that we are not claiming that establishing the heritage stage is not worthwhile, but it is not the same thing as moving towards digital vitality. And so, if we are to use the proposed framework for helping communities decide on the future of their language, it is helpful to identify a distinction between situations where a digital project has a possibility of succeeding, and those where other groundwork needs to be done first. Otherwise we risk the danger of using models which imply that a language can be revitalised by digital means alone in cases where it cannot, which breeds false hope and ultimately may discourage any efforts to expand the use of a minority language. Hence we claim that identifying the latent stage (it is possible that another name could

be chosen for this stage) is a valuable tool in the development of a framework whose goal is to encourage the appropriate activities for different patterns of established language use.

As we have noted, this framework for categorising digital use is different from scales such as EGIDS, which reflect broader use in the language community. Digital use is different from spoken use, but we must also emphasise that digital practices rely on these broader practices being sustained. In turn, a digital strategy is itself also part of a bigger picture of language use. Vigorous digital use may have a positive impact on attitudes towards the language, and on other literacy practices, and thus be part of a strategy of a minority community in maintaining the language for the longer-term future, using it as a vehicle for planning their own future and development. It is in this hope that we present this framework, to assist communities in identifying the stage they are at, and what the best next steps may be.

References

1. Abercrombie, D. (1963). *Problems and Principles in Language Study.* 2nd edition. London: Longman.

2. Brenzinger, M., Yamamoto, A., Aikawa, N., Koundiouba, D., Minasyan, A., Dwyer, A., Grinevald, C., Krauss, M., Miyaoka, O., Sakiyama, O., Smeets, R., Zepeda, O. (2003). *Language Vitality and Endangerment.* Paris: UNESCO Ad Hoc Expert Group Meeting on Endangered Languages. http://unesdoc.unesco.org/images/0018/001836/183699E.pdf.

3. Coulmas, F. (2013). *Writing and Society.* Cambridge: Cambridge University Press.

4. Crystal, D. (2008). *Txtng: the gr8 db8.* Oxford: Oxford University Press.

5. Fasold, R. (1984). *The Sociolinguistics of Society.* Oxford: Blackwell.

6. Fishman, J. A. (1991). *Reversing language shift.* Clevedon: Multilingual Matters.

7. Gibson, M. (2012a). Extinct languages and languages close to extinction: How to preserve that heritage? In: Vannini, L. & Le Crosnier. H. (eds.) (2012), pp. 75–88.

8. Gibson, M. (2012b). "The urban vernacular(s) of Nairobi: Contact language, anti-language, or hybrid language practice?" Paper presented at Sociolinguistics Symposium 19, Berlin, Germany. 21–24 August, 2012. https://www.academia.edu/1878748/The_urban_vernacular_s_

of_Nairobi_contact_language_anti-language_or_hybrid_language_ practice.

9. Githiora, C. (2002). Sheng: Peer language, Swahili dialect or emerging Creole? *Journal of African Cultural Studies* 15, 2: 159–81.

10. Kornai, A. (2013). Digital Language Death. PLoS ONE 8(10): e77056. doi:10.1371/journal.pone.0077056   http://www.plosone.org/article/ info:doi/10.1371/journal.pone.0077056.

11. Lewis, M. P. and Simons, G. F. (2011). "Ecological Perspectives on Language Endangerment: Applying the Sustainable Use Model for Language Development". Paper presented at American Association for Applied Linguistics, Chicago, 26 March. http://www.sil.org/~simonsg/ presentation/Applying%20the%20SUM.pdf.

12. Lewis, M. P., Simons, G. F. (2010). Assessing endangerment: Expanding Fishman's GIDS. *Revue Roumaine de Linguistique* 55 (2):103–120. http://www.lingv.ro/RR%202%202010%20art01Lewis.pdf.

13. Muhammed, R., Farrag, M., Elshamly, N., Abdel-Ghaffar, N. (2011). "Summary of Arabizi or Romanization: The dilemma of writing Arabic texts". Paper presented at Jīl Jadīd Conference, University of Texas at Austin, 18–19 February. https://www.utexas.edu/cola/depts/ mes/events/conferences/jiljadid2011/papers/FinalArabizSummary_ JilJadid.pdf.

14. Vannini, L., Le Crosnier, H. (eds.). (2012). *NET.LANG: Towards the multilingual cyberspace*. Paris: C&F Editions. http://net-lang.net/ lang_en.

Alfredo RONCHI

Secretary General, European Commission – MEDICI Framework of Cooperation;
Professor, Polytechnic University of Milan
(Milan, Italy)

Is the Internet a Melting Pot?

Abstract

Is the Internet a melting pot creating a new lingua franca the "Engternet"? After different waves jeopardising cultural diversity such as the different aspects of globalisation including global markets and infrastructures the Internet and related services are a potential silver bullet to kill diversities. Why a similar concern? Because once more a dominant actor comes on stage.

This aspect takes us to carefully consider the importance to preserve "diversity", especially in the digital age. What is the real value of diversity?

We all know that the world population today is bigger than the number of people that lived on the planet Earth since the human race appeared, but today it is incredibly easier to disseminate ideas and content through the planet reaching individuals.

This is one of the effects of the global inter-communication in the digital era. Moreover global software tools are unleashing everyday creativity with no regards for citizenship, language, gender or census.

On the one hand the digital age is enabling better opportunities to exploit local cultures and knowledge due to minorities, on the other hand such a "global village" jeopardizes minorities and local cultures playing the role of a standardization agent.

A kind of English language, the one generated by spelling and grammar checkers, and translators is still placed in pole position but very close we find Chinese language quickly improving its ranking. New devices and communication standards are inspiring new languages built on abbreviations, phonetic equivalences, graphic signs and emoticons, will the 140 chars tweet become the new structure of verses?

Smart phones and tablets are breaking time and space barriers including formerly divided people in the emerging cultural phenomenon. This is true both for the young generation and even for elderly people who find tablets and smart phones more user friendly than "old" computers.

Digital technology is offering new ways to express creativity in different fields: music, images, videos, physical objects and more, enabling young generation to express their feelings and contribute to the creative industries.

Introduction. Globalisation & Cultural Diversity

Years ago we all entered, willing or not, the age of globalisation. This does not only mean to drink Cuban Mojito in South Korea or enjoy Malaysian craftsmanship in Switzerland but involves deep changes in a wide range of sectors (cultural, linguistic, economic, artistic, and more). The planet has never looked so small as today, with people travelling across continents and oceans apace. The recent significant increase of travellers (even if relatively modest compared with the general population) coming from new emerging economies such as China, India and Brazil gave acceleration to such a process. On the cultural and social side there is something positive associated to globalisation: people know much more about other inhabitants of the planet, their culture, their issues and it enriches our opportunity to analyse facts, events and behaviours thanks to multiple viewpoints. This may contribute to a peaceful future. At the same time globalisation refers to dominant languages and cultures; this aspect may endanger local languages and culture. If on the one hand a global market enables multiple trading on the other hand a homogeneous language and culture simplify the business. Why to support linguistic and cultural diversity? The Universal Declaration on Cultural Diversity (UNESCO General Conference 2001) states "cultural diversity as a source of exchange, innovation and creativity is just as indispensable for humanity as biological diversity for Nature, and is a treasure shared by the entire human race". If this is not enough we can add that diversity is always a patrimony, richness, it means "life", while uniformity on the opposite sometimes means "death". Even in the creative world of moviemakers the idea of "hell" in the future is tightly connected with uniformity, absence of diversity; the Henry Ford's free choice of colour "so long as it's black"16. Today, even if for different reasons, the motto "Think different!" contributed to create the Apple community.

It is a common understanding that people who grow up in different cultures do not just think about different things, they actually think differently. The environment and culture in which people are raised affects and even determines many of their thought processes. So the Apple's "Think different!" is much more than a motto.

---

[16] *Henry Ford (model T 1908): "Any customer can have a car painted any colour that he wants, so long as it is black."*

Sometimes even intercultural initiatives such as the Erasmus programme in Europe do not really offer an opportunity to experience a different country for 100%. Thanks to Erasmus European students may experience a period of time abroad attending university courses in a different country. An Italian student may spend one semester in Spain but very often for certain reasons they do not enter in real touch with local culture and language because they use to speak in English and do not learn Spanish and adopt a "global" life-style.

The risk is to go toward a uniform language and cultural model loosing the richness due to centuries of different expressions in the field of art, literature, painting, music etc. Particularly endangered due to such a situation are "minoritized" cultures and languages.

A tight interdependency relationship between language and culture is true and evident. The grammar, the richness of vocabulary, the different forms to express a concept, the presence or absence of certain terms, simply to mention some aspects, may tell us a lot about that people. In order to fully enjoy a "culture" you must know the associated language and on the other side knowing a language you have the main entry point to the associated culture.

All the above make us conscious that linguistic and cultural diversity is the edge of an "iceberg" that includes cultural identities, sense of belonging to a community, personal root, intangible heritage, popular knowledge and achievements throughout the centuries, proper interpretation of local content and much more.

Dominant languages used in major domains such as governmental, scientific, cultural, political, economic, etc. contribute to making minority languages decline in the shadow, and together with them knowledge and cultural experience of these cultures developed through the time vanish gradually.

In order to have an idea about the size of the problem forecasts say that more than half of the currently alive 7,000 languages may extinguish within several generations. Of course the huge majority of these languages are spoken by minorities spread all-over the world.

This means that a large majority of peoples nowadays have no chance to fully express their culture and use their own language. They live in a multi-ethnic country and share the dominant culture and language, in this way most of languages are marginalized and future generations will not speak anymore

the language of their ancestors and their cultural roots will disappear in the shadow.

These aspects are so crucial for future generations that even the key documents of the World Summit on the Information Society (WSIS): the Declaration of Principles and Plan of Action (first phase in Geneva, 2003), the Tunis Commitment and Tunis Agenda for the Information Society (second phase in Tunis, 2005) and the Vision for WSIS Beyond 2015 (WSIS+10 Geneva, 2014) emphasize the importance of the preservation of cultural and linguistic diversity and suggest a set of measures necessary to achieve this goal.

"Indigenous and traditional knowledge are recognised as pathways to develop innovative processes and strategies for locally-appropriate sustainable development. This knowledge is integral to a cultural complex that also encompasses language, systems of classification, resource use practices, social interactions, ritual and spirituality. These unique ways of knowing are important facets of the world's cultural diversity, and provide a foundation for comprehensive knowledge society." Moreover "There is full respect for cultural and linguistic diversity, and for everyone's right to express themselves and to create and disseminate their work and local content in the language of their choice. The preservation of digital heritage in the information society is ensured." [Draft WSIS+10 Vision for WSIS Beyond 2015]

This set of documents, outcomes of the Summits, takes us directly to the next paragraph.

Information Communication Technologies

The previous paragraph outlines the importance to preserve and ensure cultural and linguistic diversity and the risk to jeopardize them due to globalization, but this is not enough in order to analyse the state of the art and relative treats. The recent relevant social impact due to Information Communication Technologies (ICTs) improvements makes this a turning point for cultural and linguistic diversity preservation and at the same time globalisation encourages the merge of cultures and languages into a de facto standard. The compound effect of the two factors, globalisation and ICTs, may impress a significant acceleration to the process.

This is to look at the half empty glass but, if we change the viewpoint, the digital era in which we live nowadays potentially offers new opportunities

for the preservation and preservation through promotion of linguistic and cultural diversity for equal and universal access to life-crucial knowledge.

Enabled by emerging ICTs new alphabets and languages are flourishing. As it already happened in the past for telegrams and radio amateurs, new devices and communication standards are inspiring new languages built on abbreviations, phonetic equivalences, graphic signs and emoticons. Will the 140 chars tweet become the new structure of verses?

Of course we cannot avoid considering that Internet services and information are mainly available in the dominant languages, the current absence of certain languages in cyberspace contributes to the widening of the already existing digital information gap.

It used to be said that there are more phones in Manhattan than in some developing countries; now, however, there is a shift of paradigm, and access to the network provides the discriminatory factor. This means that both a lack of physical access to the network and the inability to handle digital technologies can cause a loss of competitiveness.

Let's get a little bit into figures. According to the latest International Telecommunication Union (ITU) survey (2014) on a world population of about 7.1 billion we find 61% of people not using the Internet at all and 39% of active Internet users where the gap between developing and developed countries is 31% to 77%. If we consider the subdivision by macro-regions of the world we find in 2013, again thanks to ITU surveys, Africa – 16%, Americas – 61%, Arab States – 38%, Asia-Pacific – 32%, Commonwealth – 52% and Europe –75%.

More interesting are figures about the Internet subscription by region subdivided by fixed or mobile connections. We find in 2013 an average value of 9.8% for fixed broadband line subscribers. In the developed world this figure is 27.2% while in developing countries it is 6.1%. If we switch to wireless broadband the situation is quite different. The average value is 29.5% of which 74.8% is due to the developed world and 19.8% – to the developing world.

The presence of different languages on the web may be summarized as W3Techs.com found in 2014. They ranked the first 36 languages but we can limit our insight to the first ten.

If we consider the first ten content languages for websites as of 12 March 2014 we find:

Language	%
English	55,7%
German	6,1%
Russian	5,8%
Japanese	5,1%
Spanish	4,7%
French	4,1%
Chinese	2,9%
Portuguese	2,3%
Italian	1,9%
Polish	1,7%
All the rest	0,7%
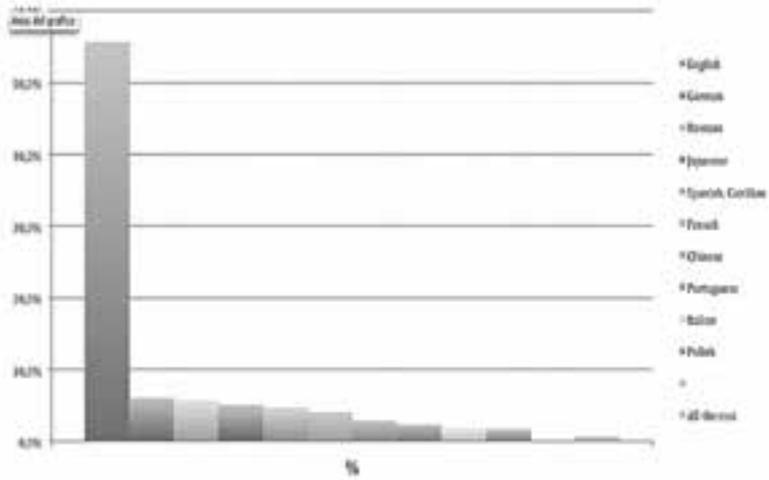
If we consider the Internet Users Languages we find:

Language	%
English	27%
Chinese	25%
Spanish	8%
Japanese	5%
Portuguese	4%
German	4%
Arabic	3%
French	3%
Russian	3%
Korean	2%
All the rest	16%
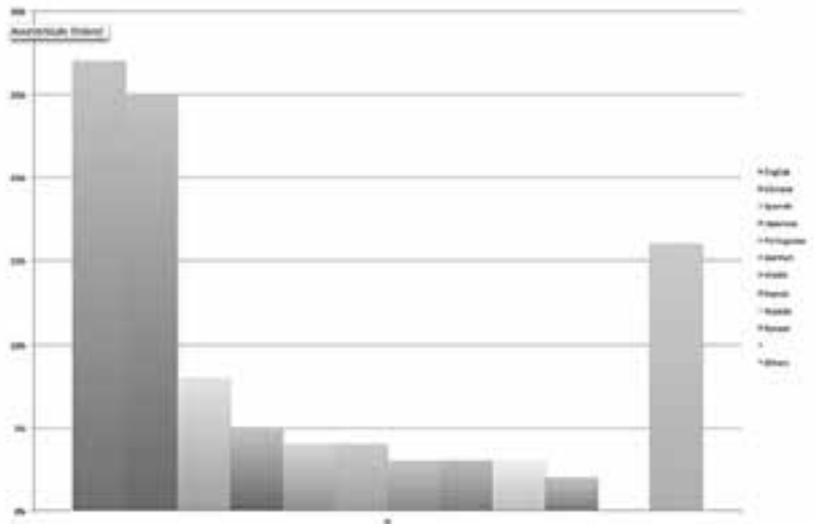
Source: "Number of Internet Users by Language", Internet World Stats, Miniwatts Marketing Group, 31 May 2011 (explanations on the methodologies used in the survey: http://w3techs.com/technologies)

Native languages are necessary instruments for social life within communities sharing the same language. They enable the expression and dissemination of social and cultural traditions, self-identification and preservation of human dignity of their speakers. As already mentioned digital technologies and tools may represent an excellent opportunity to preserve and disseminate local culture. The Internet is a powerful tool in order to preserve and disseminate cultural content, traditions and languages. The evolution of automatic online translators enabled the access to "foreign" content written in various alphabets to end-users. Thus, for instance, it is now possible to read Arabic of Chinese web pages with reasonable success. Virtual keyboards, especially on pads, provided an easy way to write in different alphabets even if addressed to relatively small communities. Other software tools or data sets such as diacritic marks spell checkers, and generally speaking natural language processors, phonetic language resources, Wikipedia, Wiktionaries will provide a significant help.

As a kind of side effect the wide diffusion of the Internet together with the social web and spelling and grammar checkers originated a kind of a new language we can term "Engternet", English on the Internet, it's a "network mutation" of the already "globalised" lingua franca.

Preservation of cultural and linguistic diversity involves relevant efforts across different countries; some countries have to deal with a number of minorities having each one a different language and culture. The general aim may resemble the protection of endangered species of animals but that's not correct. Ensuring long life to languages and cultures involves multiple efforts. Governments and international organisations cannot afford 100% of the costs and provide all the resources needed for such a mission. It is even hard to refer to the market looking for business sponsors; there is not apparently a direct return of investment apart from very well known situations or potential touristic exploitation.

One of the potential solutions is to refer to communities and crowds. Communities and crowds, these are among the most relevant resources nowadays.

It seems to be a completely new paradigm of software and services development beyond user groups and open software, the only way to face huge projects and compete with key software enterprises. The average "size" of "social" products and services is now affordable only by crowdsourcing. A number of services that do not find a proper economic dimension or even do not have the required appeal in order to be provided by companies may only rely on the "crowd". They are the potential solution to a number of problems almost impossible to be solved by business companies. How to build up a comprehensive encyclopaedia,

how to collect punctual information about the weather or traffic, how to mass digitize texts or instruct optical character reader? In the global society crowds are playing the role of "public services". Crowd sourcing offers a new paradigm in software and services development.

The idea to share something with someone else, a group of people, usually generates a sense of belonging to a "community". Communities are an integral part of history and technology; in the specific field of communication we find "amateur radio" also called *ham radio* or OM (old man) and later on the citizens' band (CB) community. Of course technical communities are not limited to the field of communications; we have computer graphics, video games, and more such as the Manga Fandom[17] but in recent times communication is the key player in the creation of communities and due to this communities directly dealing with communication means are facilitated. As already outlined social media are one of the milestones recently introduced in the digital domain. Social media is the key to success of the digital domain, the reply to the Win '95 promo "Where do you want to go today?", the real mass use of digital resources, the one creating "addiction" is the social side. Since the creation of the first blogs opening the opportunity to share opinions and beliefs with a significant number of users the number of "social" application has grown up very quickly. The evolution of online news due to the social web and the birth of "prosumers" did the rest. Twitter, YouTube, Facebook and blogs represent a real revolution in the domain of news.

However, network-based services may not be of any use to emerging countries if end-users are unable to access the information. Access to archives, cultural services, educational and training services need to be provided in *e-format* because of the added value but we must also ensure that this added value can be exploited by end-users. Emerging technologies such as *tablets*, *smart phones* and *enhanced portable communication systems* may represent a solution on the *client (application) side*. The presence of a client side does not necessary imply the corresponding presence of a *server side*; *peer-to-peer* connections offer an attractive alternative approach that enables new interpersonal services.

When dealing with cultural issues, we often face problems such as the preservation of "cultural identity" or "cultural diversity" in some technologically remote areas of the world. How do we safely store and offer oral traditions or storytelling for local public enjoyment, for instance? Steaming audio and video across the Internet requires some bandwidth in addition to the basic technology and web access, so that time ago the only way to ensure that end-

---

[17] *Manga fandom is a worldwide community of fans of Japanese cartoons manga.*

users are able to experience them was to use VHS cassettes, an "easy access" technology which was widely available, cheap and the de facto standard.

This aspect is very relevant, because if it is important to preserve cultural assets to keep records of rites, oral traditions, and performances as a legacy to humanity, we must also provide the content holders/owners with a copy of the final, released version of the "content" in an enjoyable format, as well as a percentage of any revenue obtained from it, as compensation, if there is not a "return on investment" for the "content owners", such a behaviour is known as "bio piracy".

This led us to consider another important aspect: how IPR should be managed. Communities that involve themselves in technological evolution must share information within a tailored legal framework. Intellectual property rights are an additional key point to be defined in order to avoid both the so called "bio-piracy" and road blocks on the way to digitize endangered cultural assets.

Traditionally, "copyright" and "copyleft" have been regarded as absolute opposites: the former being concerned with the strict protection of authors' rights, the latter ensuring the free circulation of ideas. In addition, with specific reference to cultural topics, the Medicean ideal to allow all mankind, regardless of social status or worth, enjoy the beauty of art seems to support free access to content.

While copyright which seeks to protect the rights of inventors to own and therefore benefit financially from the new ideas and products they originate, thus encouraging further product development is associated with a vast amount of legislation globally (leading to corresponding applicative complications), few studies have been made of copyleft. Indeed, a commonly held belief about copyleft is that it begins where the boundaries of copyright end, spreading over a no man's land of more or less illegal exploitation.

"What is worth copying is probably also worth protecting." Protecting intellectual property involves two main tasks: protecting investments and creativity, and ensuring that the moral rights to original works are assigned to the authors of those works (these are the so called "continental rights"). Preservation of endangered languages and cultures will certainly involve intellectual property issues may them be solved thanks to copyright, copyleft or other approaches such as Creative Commons.

Conclusions

To conclude I would like to introduce my experience as a member of the board of executive directors of the World Summit Award. Since 2003, thanks to my role, I have the chance to evaluate the best eContent & Services created in

more than 165 countries all over the world, the first phase of the WSIS held in Geneva. This is a unique opportunity to evaluate the state of the art of the digital "environment" in different countries, where "environment" means "readiness", infrastructure and applications. With reference to our main topic, "diversity". it is not surprising that using the same technical tools reflects the cultural background of authors. Colours, graphic, look and feel relate to the country of origin. Products coming from multi ethnic countries reflect such richness and offer a multilingual interface enabling even small communities to feel "at home".

"If you talk to a man in a language he understands, that goes to his head. If you talk to him in his language, that goes to his heart" [Nelson Mandela]

References

1. Fink, E., Ronchi, A. M. et al. (2001). On Culture in a world wide information society. MEDICI.

2. Kuzmin, E., Parshakova A. (eds.) (2012). Linguistic and Cultural Diversity in Cyberspace. Proceedings of the 2nd International Conference. Interregional Library Cooperation Centre, Moscow. ISBN 978-5-91515-048-7.

3. Ronchi, A. M. (2009). eCulture: cultural content in the digital age. Springer. ISBN 978-3-540-75273-8,

4. Vanini, L., Le Crosnier, H. (eds.) (2014). NET.LANG: Towards a Multilingual Cyberspace, C&F Editions.

SECTION 1. ICT FOR LINGUISTIC AND CULTURAL DIVERSITY IN CYBERSPACE

Mark KARAN

International Sociolinguistics Coordinator,
Senior Sociolinguistics Consultant,
SIL International
(Grand Forks, USA)

The Role of Motivational Alignment in Preserving and Developing Languages: Effective Use of Wikis, Blogs, Posts, Tweets and Text Messages

Abstract

To introduce SIL International to those not yet familiar with the organization, an overview of different ways SIL is using cyberspace to preserve and develop languages and cultures is presented. This includes online linguistics tools, dictionary creation, cataloguing the languages of the world, and readiness of languages for life in cyberspace (fonts, orthographies).

This paper then describes the Sustainable Use Model of Language Development, a comprehensive, explanatory, predictive model of language development, and then demonstrates how application of the model often reveals the need of motivational alignment within the interested speech community.

Next, the Perceived Benefit Model of language shift is described. This model identifies the motivations that lead to the community's many language choice decisions which when combined result in language shift. This then leads to a discussion of how the model's motivational analyses guides and shapes the effective use of wikis, blogs, posts, tweets and text messages in the motivational alignment needed for a sustainable multilingualism.

Introduction

SIL International is a nonprofit organization which serves language communities around the world by helping build their capacity for sustainable language development by means of research, translation, training and materials development. Two models, a language development model and a language shift model, developed within the activities of SIL, are presented in this paper in

order to demonstrate respectively how motivational alignment in a speech community is often necessary for language development, and how the speech community can bring about this needed motivational alignment.

Because of the special focus of this conference, this paper first goes on the tangent of introducing some of the different ways SIL is involved in using cyberspace to preserve and develop languages and cultures.

Preserving and Developing Languages and Cultures in Cyberspace

In addition to all SIL is doing in the area of fonts and scripts to facilitate the preservation and development of languages in cyberspace [SIL International 2014a], SIL has been adding compatibility functions to different software, allowing cyberspace collaboration in different language research and development related activities.

One example of this is an online dictionary publishing platform Webonary.org which allows members of the language community the possibility of accessing and commenting on entries of dictionaries in the development process. Here is an example from the Pacoh language of Vietnam: http://pacoh.webonary. org/. Members of the language community can search, access and comment on different entries. The comments are reviewed by the dictionary compilers, used to improve the entry, and shared on the site.

Another example is Ethnologue.com. A feedback function has been added to the website so that users can participate in improving specific language listings.

A third example is the send and receive packets on FLEx. FLEx (FieldWorks Language Explorer) is a programme for dictionary compilation, text analysis, and interlinearization. The Send/Receive Project function of FLEx supports multiple users working together on one project over the Internet.

Software and Font products can be found at http://www.sil.org/resources/ software_fonts.

The Sustainable Use Model of Language Development

The Sustainable Use Model of Language Development [Lewis and Simons 2014 pre-publication draft] is a practical, predictive, working model of how language development works and how it is best facilitated. It is structured on a new revision of Fishman's [1991] GIDS language vitality scale called the EGIDS scale [Lewis and Simons 2010] described below. It is built on the premise that local communities must be the ones making decisions concerning the future of their language, and that these decisions will be informed decisions

whereby the community members know what they must be doing in order for their choices for their language to be realized.

This model is built on the observation that there are four particular levels of vitality that are much easier for a language to stay at than all the intervening levels. These four levels are: 1. Sustainable Literacy, 2. Sustainable Orality, 3. Sustainable Identity, and 4. Sustainable History, and are described below.

This model stipulates that for a language to stay at a particular sustainable level, certain sufficient and necessary conditions must be met. These conditions are called the FAMED conditions, and are described below.

The EGIDS Scale of Language Vitality

The Expanded Graded Intergenerational Disruption Scale (EGIDS) [Lewis and Simons 2010, 2014] is a scale of language vitality based on and expanded from Joshua Fishman's Graded Intergenerational Disruption Scale (GIDS). EGIDS added some levels not on the GIDS, and split apart two of the GIDS levels where an internal distinction proved to be very important.

Level	Label	Description
0	International	The language is widely used between nations in trade, knowledge exchange, and international policy.
1	National	The language is used in education, work, mass media, and government at the national level.
2	Provincial	The language is used in education, work, mass media, and government within major administrative subdivisions of a nation.
3	Wider Communication	The language is used in work and mass media without official status to transcend language differences across a region.
4	Educational	The language is in vigorous use, with standardization and literature being sustained through a widespread system of institutionally supported education.
5	Developing	The language is in vigorous use, with literature in a standardized form being used by some though this is not yet widespread or sustainable.
6a	Vigorous	The language is used for face-to-face communication by all generations and the situation is sustainable.

6b	Threatened	The language is used for face-to-face communication within all generations, but it is losing users.
7	Shifting	The child-bearing generation can use the language among themselves, but it is not being transmitted to children.
8a	Moribund	The only remaining active users of the language are members of

the grandparent generation and older. |
8b	Nearly Extinct	The only remaining users of the language are members of the grandparent generation or older who have little opportunity to use the language.
9	Dormant	The language serves as a reminder of heritage identity for an ethnic community, but no one has more than symbolic proficiency.
10	Extinct	The language is no longer used and no one retains a sense of ethnic identity associated with the language.

Fishman's level 6 is split into levels 6a and 6b because of the importance associated with complete intergenerational transmission of the language, maintained in 6a, and absent in 6b. Fishman's level 8 is split into levels 8a and 8b because of the importance of an older generation viably using the language, maintained in 8a, and absent in 8b. Fishman's numbering order is maintained, where the higher language vitality is associated with the lower numbers, presumably because Fishman was basically talking about disruption of the language being passed from parent generation to child generation; the more disruption, the higher the number.

The EGIDS scale is now much more than a graded scale of intergenerational disruption of language. It is a good language vitality scale.

In order to determine the EGIDS level of a particular language a decision tree is used (below). Starting with the "How is the language used?" blue box on the left, if the language is used outside of its own language area, follow the arrow to the "What is the level of official use?" blue box up on the right. If the language isn't used outside of its own language area and is used as a mother tongue in homes, follow the arrow to the "What is the sustainability status?" box to the right. If the language isn't used as a mother tongue, follow the arrow to the "youngest generation" blue box on the bottom right, unless the language isn't used at all.

Then from those three big middle blue boxes, if the top statement is true, the EGIDs level is indicated. Then go down to the next highest statement, and so on.
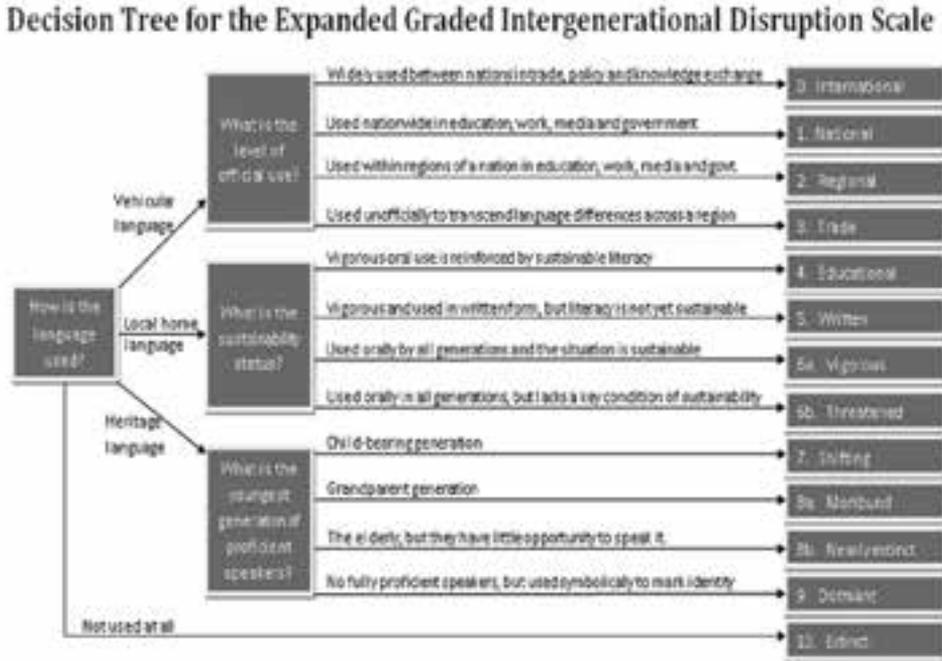
Figure 1. *Decision Tree of EGIDS Diagnostic Questions*
[Lewis and Simons 2014: 93]

Levels of Sustainable Vitality

In the SUM model, there are 4 levels of sustainable use; 3 levels of sustainable language use and 1 level of sustainable documentation. These levels are:

EGIDS Level 4 **Sustainable Literacy:**

- not only vigorous oral use but also widespread written use;
- supported (transmitted) by sustainable institutions.

EGIDS Level 6a **Sustainable Orality**:

- strong identity rooted in the language;
- vigorous oral use by all generations for day-to-day communication;
- language transmission takes place in the family or local community.

EGIDS Level 9 **Sustainable Identity**:

- no fully proficient speakers;

- a community associates its identity with the language;
- not used for day-to-day communication; used ceremonially or symbolically.

EGIDS Level 10 **Sustainable History** (level of sustainable documentation):
- no remaining speakers;
- no one associates their identity with the language;
- a permanent record (history) of the language is preserved.

The following graphic illustrates the levels of sustainable language used as plateaus on a slope representing language vitality levels:
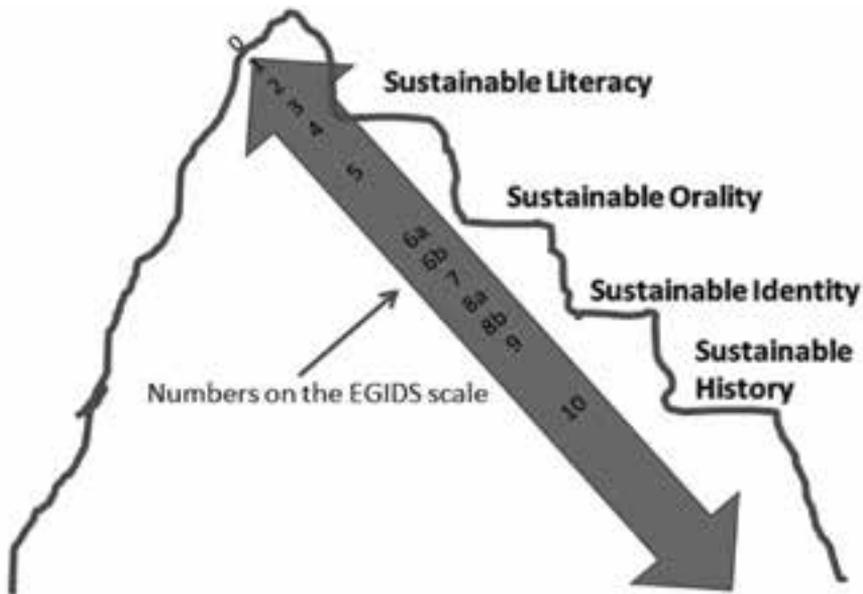
The important premise about sustainable levels is that all other levels, without intervention, will naturally decay to the next lower level of use. Once a language goes over the edge of the sustainable plateau, it is on the steep slippery slope to the next lower sustainable level.

The FAMED Conditions

In order for a language to stay at a Sustainable Level, five conditions must be met; the sufficient and necessary FAMED Conditions. All five conditions are essential for the sustainable vitality level to be maintained. The FAMED acronym is:

Functions

Acquisition

Motivation

Environment

Differentiation

- **Functions** – Deals with how the language is useful and used by the community.

- **Acquisition** – Deals with people learning the language.

- **Motivation** – Deals with the motivations of the community members to use the language.

- **Environment** – Deals with the external environment (e.g., majority group attitudes toward the language).

- **Differentiation** – Deals with societal norms for regularly using the language in specific domains.

Or expressed differently:

- **"Functions** – Functions (uses, bodies of knowledge) associated with the language must exist and be recognized by the community.

- **Acquisition** – A means of acquiring the needed proficiency to use the language for those functions must be in place and accessible to community members.

- **Motivation** – Community members must be motivated to use the language for those functions. They must perceive that the use of the language is beneficial in some way.

- **Environment** – The external environment (e.g. national, regional, or local policy) must not be hostile to the use of the language for those functions.

- **Differentiation** – Societal norms must clearly delineate the functions assigned to the local language marking them as distinct from the functions for other languages in the speech community's repertoire." [Lewis and Simons 2014: 127]

The following chart [Simons and Lewis 2012] (provided as a handout: SUM at a Glance) presents the FAMED conditions for EGIDS levels 4, 5, and 6a.

EGIDS Level	Functions	Acquisition	Motivation	Environment	Differentiation
4: Educational (Sustainable Literacy)	Adequate vernacular literature exists in every domain for which vernacular writing is desired.	Vernacular literacy is being taught by trained teachers under the auspices of a sustainable institution.	Members of the language community perceive the economic, social, religious, and identificational benefits of reading and writing in the local language.	Official government policy calls for the cultivation of this language and cultural identity and the government has put this policy into practice by sanctioning an official orthography and using its educational institutions to transmit local language literacy.	Members of the language community have a set of shared norms as to when to use the local language orally and in writing versus when to use a more dominant language.
5: Written (Incipient Literacy)	Enough literature exists in some domains to exemplify the value of vernacular literacy.	There are adequate materials to support vernacular literacy instruction and some members of the community are successfully using them to teach others to read and write the language.	Some members of the language community perceive the benefits of reading and writing their local language, but the majority of them still do not.	Official government policy encourages the development of this language. Official government policy has nothing to say about ethnolinguistic diversity or language development and thus raises no impediment to the use and development of this language.	Members of the language community have a set of shared norms as to when to use the local language orally versus when to use a more dominant language, but for writing, some members of the language community use the local language in written form for particular functions while others use a more dominant language for many of the same functions.

EGIDS Level	Functions	Acquisition	Motivation	Environment	Differentiation
6a: Vigorous (Sustainable Orality)	Adequate oral use exists in every domain for which oral use is desired (but there is no written use).	There is full oral transmission of the vernacular language to all children in the home (literacy acquisition, if any, is in the second language).	Members of the language community perceive the economic, social, religious, and identificational benefits of using their language orally, but they perceive no benefits in reading and writing it.	Official government policy affirms the oral use of the language, but calls for this language to be left in its current state and not developed.	Members of the language community have a set of shared norms as to when to use the local language orally versus when to use a more dominant language, but they never use the local language in written form.

Source: Simons and Lewis 2012

From this chart, one can see how the particular FAMED Conditions are different for each condition and for each vitality level. Note, for example, the differences between the Motivations for Level 4 Educational and the Motivations for Level 6a Vigorous. And again, In order for a language to stay at a Sustainable Level, the FAMED Conditions for that level must be sustained. All five conditions are essential for the vitality level (EGIDS) to exist. And in order for a language to get to a higher level, all five of the FAMED Conditions for that higher level need to be met.

Using the SUM

The process of using the SUM involves first identifying the speech community where it is to be applied. It is important to note that the speech community, and not the language community, is the appropriate level on which the SUM is to be applied. To quickly differentiate the two, the Ewe language community includes the Ewe speakers in Accra and other cities in Ghana, the rural dwelling Ewe speakers, and the Ewe speakers in diaspora, living in England for example. This Ewe language community is composed of at least the three Ewe speech communities mentioned above, those in the cities, those in rural areas, and those living in England. A speech community is basically a group who sees themselves as a group and shares a language repertoire and language use norms.

Having identified the speech community in focus, the first step is to facilitate them in doing an EGIDS analysis for their language. Then, after they have been familiarized with the concept of sustainable and non-sustainable levels, the next step is for the speech community, at a culturally appropriate in-group meeting, to select the sustainable level it desires to be at. The third step then is that the group is facilitated in doing a FAMED analysis of what their actual language vitality level is.

This can be graphically noted on a SUM chart like the following, with the red shapes indicating the desired vitality level and the yellow shapes indicating the actual vitality level. In this charted example, the community desires to be at Level 6a Vigorous, the Sustainable Orality level. Their actual FAMED analysis has shown that they are on that level for Functions, Environment, and Differentiation, but that they are only on the 6b Threatened level for Acquisition and Motivation. This indicates that in order for them to get to and stay at the 6a Vigorous, Sustainable Orality level, they need to see a change in their community acquisition and motivation profiles so that the actual situation matches the FAMED Conditions of the desired sustainable level. The facilitators then can share with the group what activities have been successfully used in other situations around the world to bring about the needed step-ups, the changes in the actual FAMED condition needed to match the FAMED conditions of the desired level.

Level of Use*	Functions	Acquisition	Motivation	Environment	Differentiation
4: Educational (Sustainable Literacy)	1. Adequate vernacular literature exists in every domain for which vernacular writing is desired.	1. Vernacular literacy is being taught by trained teachers under the auspices of a sustainable institution.	1. Members of the language community perceive the economic, social, religious, and identificational benefits of reading and writing in the local language.	1. Official government policy calls for the cultivation of this language and cultural identity and the government has put this policy into practice by sanctioning an official orthography and using its educational institutions to transmit local language literacy.	1. Members of the language community have a set of shared norms as to when to use the local language orally and in writing versus when to use a more dominant language.
5: Written (Incipient Literacy)	1. Enough literature exists in some domains to exemplify the value of vernacular literacy.	1. There are adequate materials to support vernacular literacy instruction and some members of the community are successfully using them to teach others to read and write the language.	1. Some members of the language community perceive the benefits of reading and writing their local language, but the majority still do not.	1. Official government policy encourages the development of this language. 2. Official government policy has nothing to say about ethnolinguistic diversity or language development and thus raises no impediment to the use and development of this language.	1. Members of the language community have a set of shared norms as to when to use the local language orally versus when to use a more dominant language, but for writing, some members of the language community use the local language in written form for particular functions while others use a more dominant language for many of the same functions.
6a: Vigorous (Sustainable Orality)	1. Adequate oral use exists in every oral use domain for which oral use is desired (but there is no written use).	1. There is full oral transmission of the vernacular language to all children in the home (literacy acquisition, if any, is in the second language).	1. Members of the language community perceive the economic, social, religious, and identificational benefits of using their local language orally, but they perceive no benefits in reading and writing it.	1. Official government policy affirms the oral use of the language, but calls for this language to be left in its current state and not developed.	1. Members of the language community have a set of shared norms as to when to use the local language orally versus when to use a more dominant language, but they never use the local language in written form.

Level of Use*	Functions	Acquisition	Motivation	Environment	Differentiation
6b: Threatened	1. Adequate oral use exists for some domains for which oral use is desired (but not for all).	1. The language is used orally within all generations but only some of the child-bearing generation are transmitting it to their children in the home.	1. Members of the child-bearing generation perceive the benefit of using their language orally for some purposes, but for others find more benefit in shifting to a more dominant language.	1. (as above)	1. Some members of the child-bearing generation use the local language orally only for functions that were traditionally reserved for the local language, while others use a more dominant language for many of the same functions.
7: Shifting 8a: Moribund 8b: Nearly extinct	1. There are entire generations that no longer have full oral use of the language.	5. The only transmission of the language is for identificational use (often in institutional settings rather than the home).	1. The child-bearing generation finds no practical benefit in speaking the language, though they may still find sentimental benefit.		
9: Dormant (Sustainable Identity)	1. Enough oral use exists to symbolize the identity of the group (but not for full communication). 2. Adequate documentation of the language exists so that revitalization would be possible.	1. (as above)	1. Members of the language community have a strong sentimental attachment to their language, but are no longer able to speak it regularly.	1. (as above)	1. The only remaining domain of local language use is identificational.

Different types of activities are necessary to address different needed step-ups. For example, if the needed step up was in the Function condition and had to do with literacy, the activities could be materials preparation. If the needed step-up was in the Environment condition, the activities could be external advocacy to change the Environmental situation. If the needed step-up was in the Motivations, Acquisition, or Differentiation condition area, the appropriate activities could be internal advocacy, to change the community's Motivation, Acquisition or Differentiation patterns so that the new norms match the desired FAMED profile levels.

A group can modify its chosen sustainable level at this time. If the group see that they don't have the will or ability to bring about the needed changes, the step ups where the actual FAMED conditions don't match the FAMED conditions of the desired level, they can choose a lower sustainable level and prepare themselves for the realities of being at that lower level.

Motivational Alignment

The comparison of a speech community's actual FAMED level with their desired FAMED level will often reveal where there are differences in language related motivations between members of the speech community. For example, where some parent-aged members of the speech community think it is best to raise their children in the language of wider communication and others think it is best to raise them in first the mother tongue and then later in both the mother tongue and the language of wider communication. These circumstances would result in a 6b Threatened status for Motivations and Acquisition in the actual FAMED analysis. This is a situation that will bring about language loss as level 6a Vigorous is needed in order for the speech community to stay at the Sustainable Orality level. Motivational Alignment then has to do with the actions of speech community members, in the interest of sustainable language use, using internal advocacy to attempt to change the motivational patterns of those in their community so that the motivations that would lead to the decline of the language are changed.

It is actually quite common that a comparison between a community's actual and desired FAMED Conditions will reveal a needed step-up in the area of Motivations, and in the areas quickly affected by the Motivations; Functions, Acquisition, and Differentiation. In the past, many of these motivations related situations were addressed with literature production activities. For example: not all of our people are teaching the language to their children, let's produce a dictionary and grammar and some stories. These activities rarely achieve their purpose. It is better to try to solve motivational issues with motivational solutions, not with literature. The best way to address motivation related needs

is with internal advocacy; part of the group reaching out to the rest with good and persuasive arguments.

The Sustainable Use Model of language development would then suggest internal advocacy actions as solutions to the cases where the Motivations in the FAMED analysis didn't match up with the Motivations needed for arriving at or staying at the desired FAMED and EGIDS level.

The Perceived Benefit Model of Language Shift

Karan [2001] presented the Perceived Benefit Model of Language Shift (and Change). In this explanatory model of language shift (and change) the concept of motivations is central. Individuals choose the languages, dialects, and styles that they think will bring them the most perceived benefit. Thus, change and shift are explained by individuals' choices. People choose to use language, dialects and styles that they think will do them good. They also make motivated choices to acquire those languages, dialects and styles that are of benefit to them. Shift in the speech community is seen as the conglomerate of individually motivated choices. People are seeking what they perceive to be for their good or for the good of their offspring, and make choices. This Perceived Benefit Model is based on the works of Bourdieu [1982], Coulmas [1992], and Labov [1965]. It involves a certain economy of languages where shift is motivated and can be seen in a synchronic cross section of the population through variation studies.

An important concept of the Perceived Benefit Model is that the motivations behind the many individual decisions that constitute language shift could be listed in a limited taxonomy of motivations. Karan [2011: 143] identifies these motivations as:
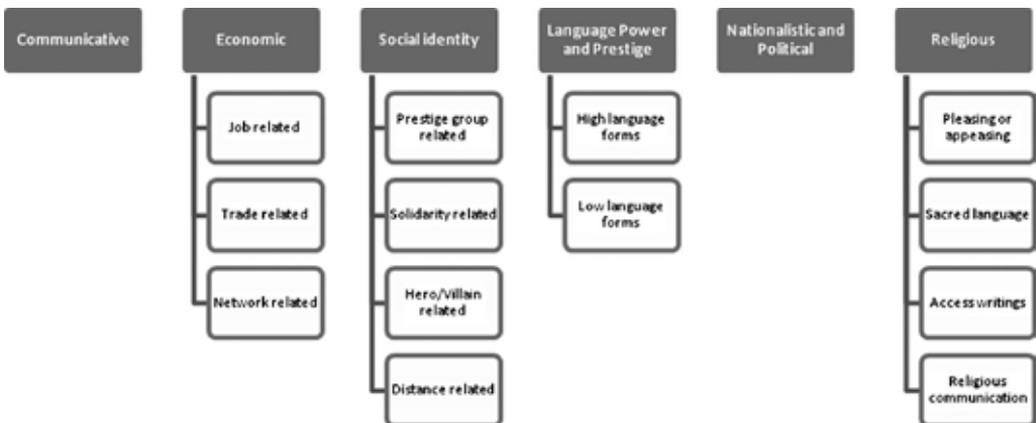
When someone is making a choice of a language, dialect, or even style of language to use, it is most likely motivated by one of these considerations.

Certain motivations are most commonly seen with official languages and languages of wider communication, while other motivations are most commonly seen with smaller and minoritized languages. Economic and Social Prestige motivations are often those behind choices for the larger, higher status languages, while group Solidarity and Identity are often those behind choices for the smaller, lower status languages. Karan and Corbett (in press) demonstrate the importance on Identity and Affiliation in decisions to maintain or use smaller, lower status languages.

The Perceived Benefit Model and Motivational Alignment

Application of the Perceived Benefit Model often included motivational studies, to determine what motivations are behind the choices for what languages. In the context of motivational alignment, where a part of the group is reaching out to the rest of the group with good and persuasive arguments to change motivations that would lead to the decline of a language, knowledge of what motivations are associated with what languages is vital. In the internal advocacy of motivational alignment, if the group desiring the change is aware that the typical motivations leading to the use of the smaller language are Social Identity, Group Affiliation, and Social Solidarity, they will most likely use those motivations when trying to influence the others to motivationally align with them. Internal advocacy motivational alignment is most effective when it is focusing on the motivations that already exist in those who have the desired motivations. These motivations are those that are the most likely to influence and convince those who are being addressed.

Motivational Advocacy in Cyberspace

Cyberspace is increasingly becoming a more and more used and effective medium of communication. It is especially effective in areas relating to internal advocacy and motivational alignment (motivational advocacy) because it is seen as real people communicating with real people, where radio, television and typical print media is more seen as the establishment talking to the people. Thus wikis, blogs, posts, tweets and text messages are vital choice channels of communication when dealing with this and other areas of speech community in-group communication.

When the motivational analysis of the Perceived Benefit Model indicates what motivations are to be appealed to for the needed motivational advocacy, and this appeal is well made through the use of wikis, blogs, posts, tweets and

text messages it can be very effective in achieving the motivational alignment needed for a sustainable multilingualism.

Another area in which the Perceived Benefit Model can be of help in this area is in the choice of the people or personas involved in the needed motivational alignment. The Perceived Benefit Model has shown that people want to be like, and emulate, the people they respect, the people they admire, the people they want to become more like, the people they want to associate with. Thus with wikis, blogs, posts, tweets and text messages it is important that the authors be people or personas that fit this profile.

If one of the top football players in a country tweets that he and his wife are raising their child in the local language, and then that gets shared in blogs and posts and text messages, it can be incredibly effective. If the population would respect and want to be like a rich, good looking lawyer and doctor couple living with their two wonderful children in a spacious villa with two luxury cars in the circle driveway, behind remote controlled security gate, that type of couple should be the preferred author of the blogs and posts advocating for the desired language motivations and use.

It is, of course, the case that this type of control of author is impossible in cyberspace. Everybody there is an author. It is however a good concept to keep in mind for where there is some available choice in introducing ideas and choosing people to officially champion ideas and campaigns.

Conclusion

In language development processes, the Sustainable Use Model can be very helpful in identifying what actions need to be taken and by whom in order to achieve the desired results. And these actions often have to do with motivational advocacy done by insiders to influence other insiders to adopt those motivations and language use patterns that will facilitate the language remaining at or arriving at the desired sustainable level. The use of the model shows how often the needed response is not a publication of a book, but rather advocacy with the speech community.

The Perceived Benefit Model can be very helpful in indicating what motivations to call upon for needed internal motivational advocacy. It can also be helpful in suggesting what people or personas are best as spokespeople for needed advocacy, as people emulate those they desire to be like.

Cyberspace, being seen as real people talking to real people is often an ideal media for the in-group communication intrinsic to the motivational advocacy needed for sustainable language development.

These aspects of knowing what to do, how to do it, with whom to work, and using what media, are very valuable in achieving language development goals.

References

1. Bourdieu, P. (1982). Ce que parler veut dire: L'économie des échanges linguistiques. Paris: Fayard.

2. Coulmas, F. (1992). Language and economy. Oxford: Blackwell.

3. Fishman, J. A. (1991). Reversing language shift. Clevedon, UK. Multilingual Matters Ltd.

4. Karan, M. E. (2001). The dynamics of Sango language spread. Dallas: SIL International.

5. Karan, M. E. (2011). Understanding and forecasting ethnolinguistic vitality. Journal of Multilingual and Multicultural Development 32(2):137–149.

6. Karan, M. E., Corbett, K. M. (forthcoming). To appear. The Importance of Identity and Affiliation in Dialect Standardization. In: Carrie Dyck, Tania Granadillo, Keren Rice, and Jorge Emilio Rosés Labrada, (eds.), Papers from the workshop on dialect standardization. Cambridge: Cambridge Scholars Publishing.

7. Labov, W. (1965). On the mechanism of linguistic change. Charles W. Kreidler (Ed.), Georgetown University Monograph Series on Languages and Linguistics, No. 18: 91-114. (Reprinted in J. Gumperz and D. Hymes (Eds.), Directions in sociolinguistics, 1972, Pp. 512–538. Also in Labov, William, 1972. Sociolinguistic patterns).

8. Lewis, M. P. (ed.). (2009). Ethnologue: Languages of the world. 16th Ed. Dallas: SIL International.

9. Lewis, M. P., Simons, G. F. (2010). Assessing endangerment: Expanding Fishman's GIDS. Revue Roumaine de Linguistique 55(2): 103–120. http://www.lingv.ro/RRL%202%202010%20art01Lewis.pdf.

10. SIL International (2014a). Web site. http://www.sil.org/resources/software_fonts.

11. Simons, G. F., Lewis, M. P. (2012). SUM at a Glance: Sustainable Use Model for Language Development. Handout/Teaching Aid.

Marcel DIKI-KIDIRI

Former Senior Researcher,
Language, Languages and Black African Cultures Unit,
French National Centre for Scientific Research
(Bangui, Central African Republic)

Terminology as a Key Step in the Promotion of Languages

Introduction

Although each step on the way of promoting any given language is important, such as elaborating an orthography in a writing system where there is not yet a standardized one for that language, writing grammars, manuals and dictionaries, teaching it in schools, and so on, only the elaboration of appropriate and standardized terminologies can allow its use in a much wider range of new specialized domains.

As an example, let us look at the use of Sängö language to deliver security announcements on board of KARINOU Airlines. Sängö is the official and national language of the Central African Republic along with French as the other official language. Although Sängö is widely spoken in the country, it lacks standardized vocabularies for a large range of special domains, and this includes aircraft flights.

KARINOU Airlines is a private Central African company that wants to use Sängö for its security and commercial announcements on board. But how can it be done? Let us suggest looking at a sample of a typical bilingual announcement in French and English from Air France documentation. We would like to apply our cultural terminology method to try and translate this announcement text into Sängö. First of all, we shall underline all words and phrases which would need some kind of treatment or explanation before being correctly translated. Then, we shall analyse each of them in their context of use in order to find out the best way to word them in Sängö taking into account both linguistic and cultural representations of their concepts. This helps looking for best Sängö equivalents as we translate the announcement into that language. The output Sängö wordings are then used in sentences to check whether they can be easily and smoothly used in a fluent speech. Only then a terminology wordlist is generated for further use as a reference.

1. Translation and Comment of a Typical Security Announcement

As the source text is both in French and English, we respect the order in which the sentences are performed in the video. That is why sometimes French comes first and some other times English comes first. This doesn't affect the Sängö translation.

1. Madame Monsieur, bonjour et bienvenue à bord.

Welcome on board, Ladies and Gentlemen.

Yäpakara na Pakara, nzönî gängö na yângö.

The phrase «à bord» or «on board» comes from the terminology of boat navigation. As a matter of fact, aviation vocabulary has been mainly borrowed from boat navigation. The original Sängö people were and largely still are riverside canoe navigators. So, naturally, the best equivalent of the phrase "on board" is *na yângö* which literally means "inside a canoe".

2. *For your safety and comfort, please take a moment to watch the following safety <u>video.</u>*

Ce <u>film</u> concerne votre sécurité à bord. Merci de nous accorder votre attention.

<u>Sindimäa</u> sô ayeke fa na âlalêgë tî dutï na sîrîrî kwê na yângö. Nzönî, âlamûkêtêtângo tî bâa nî sï.

We may notice here that «video is used in English while "film" is preferred in French. So, in Sängö, the best equivalent is "sindimäa", movie.

3. Chaque fois que ce <u>signal</u> est allumé, vous devez <u>attacher</u> votre <u>ceinture</u> pour votre <u>sécurité</u>. Nous vous recommandons de la maintenir attaché de façon visible lorsque vous êtes à votre siège.

Whenever the <u>seat belt sign</u> is on, your <u>seat belt</u> must be <u>securely fastened</u>. For your <u>safety</u>, we recommend that you keep the seatbelt under visibility all the time you are seated.

Töngana <u>wâfâ</u> sô azä, kânga<u>darakûba tî kitï</u> tî mo, tî <u>bataterê</u> tî mo. Nzönî mo zîa nî dandaranatângo sô kwê mo ngbâ tî dutï.

It appears that in English there is a need for saying things much more precisely or much more in detail than in French. Thus, the French "signal" is reflected by "seat belt sign", "ceinture" by "seat belt" and "attacher" by "securely fastened". We should also notice that in French, the phrase "pour votre sécurité" ends the first sentence whereas in English, its equivalent "For your security" starts the second sentence. In Sängö, we use "wâfâ" which means "light signal" for the equivalent of "signal" in this context, and "darakûba tî kitî" which means "belt

of seat" for translating "seat belt". It is worth pointing out that both "darakûba" ("belt") and "kitî" ("armchair"), are old words which are not commonly used by young generations of Sängö speakers. So, they are available to be recycled into a technical use. Finally, we end the first sentence with "tibataterê tî mo" (literally: "to protect body of you") which is the equivalent of "for your safety / pour votre sécurité".

4. To <u>release</u> the seat belt, just lift the <u>buckle</u>.

Pour <u>détacher</u> votre ceinture, soulevez la partie supérieure de la <u>boucle</u>.

Tî <u>zâra</u>darakûba nî, yâandöbê tî <u>bïngï</u> nî na ndüzü.

The Sängö verb "zâra" reflects more the English "to release" than the French "detacher" ("untie"). We use the word "bïngï" which means "ring" to translate "buckle", a short way of saying "belt ring", since "darakûba" ("belt") is already mentioned in the same sentence.

5. <u>Il est strictement interdit</u> de fumer dans l'avion, y compris dans les toilettes.

This is a no smoking flight. And <u>it is strictly prohibited</u> to smoke in the toilets.

<u>A ke kâsâ kâsâtîtene</u>zo anyön mânga na y â tî lapärä sô, ngâ na yâ tî kabinïi.

There is no real difficulty of translation in this fifth example. We simply point out that in English the verb "prohibit" is preferred to the verb "forbid" in this context. We therefore make sure that in Sängö the prohibition is clearly understood. The expression "ake kâsâ kâsâ" reflects that strong will to strictly prohibit smoking in the plane.

6. En cas de <u>dépressurisation</u>, un <u>masque à oxygène</u> tombera automatiquement à votre portée.

If there is a sudden <u>decrease</u> in the <u>cabin pressure</u>, your <u>oxygen mask </u>will automatically drop in front of you.

Töngana <u>pëtëpupu</u> tî yâ tî lapärä nî <u>atîa</u>, fade <u>tagï tî tâsôkö</u> atï lo ôkonagbelê tî mo.

This is an example of how using several source languages can help finding the best approach to translation. While the French word "dépressurisation" sounds very technical and difficult to translate into a language where the concept of air pressure is not commonly known, the English wording "decrease in the cabin pressure" is more explicit hence giving to the translator a better way to express the same idea in Sängö such as "If the air pressure happens to lack...". Air pressure is translated by an easy-to-understand neologism "pëtëpupu", pressure (of) air. The verb "tîa" means "to lack, to miss" and stands for "to decrease". That is what happens when there is no more enough air in the plane.

7. Tirez sur le masque pour <u>libérer l'oxygène</u>. Placez-le sur votre visage.

Pull the mask toward you to <u>start the flow of oxygen</u>. Place the mask on your nose and mouth.

Gbôto kâmba tî tagï nî sï <u>tâsôkö nî asua.</u> Leke tagï nî na ndö tî hôn tî mo na yângâ tî mo.

Once more, the French language is more synthetic with "libérer" where the English language is more analytic and explicit by saying "start the flow of". Sängö is closer to the English wording as it says "sï tâsôkö nî asua" which means literally "then the oxygen flows".

8. *Make sure your own mask is <u>well adjusted</u> before helping others.*

Une fois votre masque <u>ajusté</u>, il vous sera possible d'aider d'autres personnes.

Töngana mo <u>leke</u> tî mo tagï nî <u>mbîrîmbîrî</u> awe, mo lîngbi tî mû mabôko na mbênî zo.

It is clear that, the wording in a given language may be different in another, provided the meaning of the message remains the same and reliable. In this example 8, Sängö says "leke (…) mbîrîmbîrî", literally to fix well, where English and French use one word "adjusted /ajusté".

9. <u>En cas d'évacuation</u>, des <u>panneaux lumineux EXIT</u> vous permettent de localiser les <u>issues de secours</u>. Repérez maintenant le panneau EXIT le plus proche de votre siège. Il peut se trouver derrière vous.

<u>In case of an emergency</u>, the <u>illuminated EXIT signs</u> will help you locate the <u>exit doors</u>. Please, take a moment now to locate the exit nearest to you. The nearest exit may be behind you.

<u>Nagbâgbûru,</u> sô zo kwê <u>adu tî sïgî,</u> fadë <u>zängö wâfâ EXIT</u> afa na mo <u>yângâda tî sïgî daä</u>. Bâa mbîrîmbîrî yângâda wa laâ ayeke ndurü na mo. Alîngbi tî dutï lo sô na pekô tî mo.

In this example, the French word "evacuation" does not explicitly refer to the urgent side of the situation as the English word "emergency" does. Yet, we chose to reflect both aspects of emergency and evacuation in the first phrase: "Na gbâgbûru" in case of emergency, "sô zo kwê adu tî sïgî", when everybody must go out.

The French "panneaux lumineux EXIT" is rendered in English by "illuminated EXIT signs" and in Sängö by "zängö wâfâ EXIT", literally "lightening light sign". The translation of "issue de secours" in French or "exit doors" into Sängö is not difficult. Yet, the Sängö wording "yângâda tî sïgî daä" literally means

"doors to go out through". As it is worded, it is not possible to work out a Sängö technical term just like in French "issue de secours" or in English "exit doors".

10. Pour <u>évacuer l'avion</u>, suivez le <u>marquage lumineux</u>.

In event of evacuation, pathway enlightened on the floor will guide you to the exits.

<u>Tî sïgî nasûkpê</u>, <u>mûgïlêgë tî wâ</u> sô azä na sêse.

In the above example, the use of "évacuer" in French and "evacuation" in English both imply the idea of the emergency conditions of getting out of the plane.

11. Les portes seront ouvertes par <u>l'équipage.</u>

Doors will be opened by the <u>cabin crew.</u>

<u>Âwakua tî yâ tî lapärä</u> nî laâ ayeke zîâyangâda nî.

How to translate "equipage" or "cabin crew" as in Sängö there is no single word for this concept? We decided to be very straightforward by saying "âwakua tî yâ ti lapärä nî", literally "workers of inside the plane". It is verbose but it is immediately understood.

12. Les <u>toboggans se déploient</u> automatiquement.

The <u>emergency slides</u> will automatically <u>inflate</u>.

<u>Ângözënë</u> ayeke <u>vulangagïâla</u> ôko.

The word "toboggan" as well as the original object it refers to come both from American Indians' culture and language. It has spread into both French and English languages. We could have borrowed it likewise in Sängö. But in this emergency context, it seems better to follow the English example which has chosen to say "emergency slides" instead of "toboggan" which is somehow associated with children games. Therefore, we coined the Sängö neologism "ngözënë" which means "canoe for sliding in".

13. Le <u>gilet de sauvetage</u> est situé sous votre <u>siège</u> ou dans <u>l'accoudoir</u> central.

Your <u>life jacket is</u> under your <u>seat</u> or in the central <u>armrest</u>.

<u>Kangalïngäbatafî</u>tî mo ayeke nagbe tî <u>kitîwala</u> na yâ tî <u>wotï</u> tî bê nî.

The French concept expressed by "gilet de sauvetage" puts forward the idea of rescue while the English "life jacket" highlights the *life* of the rescued person. This helps understanding that the real function of the jacket here is to protect the life of the person who wears it. So it becomes easy to translate this term by "kangalïngäbatafî" literally: jacket life_protector.

In a plane, seats are actually armchairs, but it is the general term "seat" that is commonly used. In Sängö, the word "kitî" originally refers to a long armchair in which middle aged people rest. We use it as a technical term for "a seat in a plane". Once more, it is easier to translate from the English "armrest", in Sängö "wotï" ("rest_arm"), than from the French "accoudoir" built upon the word "coude" ("elbow").

14. Passez la tête dans l'encolure, attachez et serrez les sangles.

Place it over your head and pull the straps tightly around your waist.

Yôro li tî mo na yâ tî <u>dû tî gônî</u>, mo gbôto<u>âkâmba</u>nî ngangü, mo kânga na kate tî mo.

In French, you need to "pass your head through the collar hole of the life jacket", while in English it is the jacket you need to pass over your head! This time, Sängö is just like French.

Although it is quite accurate to translate "sangles" or "straps" by "kâmba" ("cord") which is a generic term, it is worth noticing that the Sängö language doesn't provide words for specific kind of cords such as mentioned above.

15. *<u>Inflate</u> your life jacket by pulling on the <u>red tackles</u>. Do this only when you are outside of the aircraft.*

Une fois à l'extérieur de l'avion, <u>gonflez</u> votre gilet en tirant sur les <u>poignets rouges</u>.

Kü mo sï na gîgî kwê awe sï mo gbôto<u>âbengbä</u><u>ligbô</u>nî tî <u>to pupu</u>na yâ tî kangalïngä nî.

The French expression «poignets rouges» and the English «red tackles» are accurately and literally rendered by «âbengbä ligbô" in Sängö. The Sängö word "ligbô" means literally "handler".

16. Nous allons bientôt <u>décoller</u>. La <u>tablette</u> doit être rangée, et votre <u>dossier redressé</u>.

In preparation of <u>take-off</u>, please, make sure that your <u>tray table</u> is <u>stowed in secure,</u> and your <u>seat back</u> is in an <u>upright position</u>.

Ë gä ndurü tî <u>löndö</u> awe. Kânga<u>kêtêmêzä</u>tî gbelê tî mo daä. Gbôto <u>bëkë tî kitî</u>tî mo <u>alütï</u>.

In French, a plane is looked at as something that is stuck on the soil. So, when it takes off, French says it "unsticks" (décoler). In Sängö, a plane "gets up" (löndö). In English, you have to "stow in secure" a "tray table". In French, you want to "put in order" a "small table". In Sängö, you "close... up" (kanga...

daä) the "small table" (kêtêmêzä) in front of you. A very short expression in French, "dossier redresé" has a more verbose version in English "seat back in an upright position". The Sängö reflex says: pull the "back of your armchair" till it "stands upright". As we can see, each language always allows slightly different representations of the same ideas, and this diversity of perception coins the different ways of expressing the same idea.

17. L'usage des <u>appareils électroniques</u> est interdit pendant le décollage et <u>l'atterrissage</u>.

The use of <u>electronic devices</u> is prohibited during take-off and <u>landing</u>.

A ke kâsâtîtene zo azä<u>âfonôno tî dadä</u>nandembë sô lapärä nî ayeke löndöwala ayeke <u>zunda</u>

At a first glance on this sentence, you may wonder how shall we translate "electronic devices / appareils électroniques" in Sängö? Indeed, the Central African traditional culture doesn't know about these things, but in the modern society everybody is used to radio broadcasting and all kinds of audio-visual sets. In Sängö these are called "fonônö". On the same time, the word "dada" which basically refers to a certain quantity of electric power, is now more and more used to mean electronic power. Henceforth, putting the two words together in a noun phrase such as "fonônö tî dada" (sets using electronic power) provides a good equivalent for "electronic devices / appareils électroniques".

Although the Sängö verb "zunda" correctly translates the English "to land" or the French "atterrir", it is interesting to point out that the meaning of the Sängö verb "zunda" doesn't include any connotation of "land". It actually describes the falling of a leaf that goes down smoothly in the air regardless of whatever it falls on. So, the conceptualization of the movement of the plane going down to land is built from a slightly different cultural point of view.

18. Les <u>téléphones portables</u> doivent rester <u>éteints</u> pendant tout le <u>vol</u>.

<u>Mobile phones</u> must remain <u>switched off</u> for the duration of the <u>flight</u>.

Fôko mo <u>mîngosînga tî bozo</u> kwê natângo sô kwê lapärä nî angbâ tî <u>huru</u>.

It is quite interesting to elicit the different cultural points of view released by the wording of the technical terms in the above examples. A "mobile phone" is a phone that you can carry on everywhere you go, therefore it is called in French "telephone portable". As it is usually carried in a pocket, it is called in Sängö "sînga tî bozö" (pocket phone). This is one of the best evidence of what in cultural terminology approach we call the diversity of the observation of the reality.

The same way, the English language says "switch on / off" whereas both French and Sängö use the metaphore of "light" by saying "allumer / éteindre" and "zä / mîngo" respectively, which mean "to light / to extinct".

And a last comment here, let us notice that in English and in French, the nouns "flight" and "vol" are commonly used in the context of the above examples. But in Sängö, it is the verb "huru" (to fly), that is convenient here because the noun "hürü" which is strictly the reflex of "flight / vol" is a neologism not yet commonly used. So the announcement is much more immediately understood if we say "during all the time the plane continues to fly" rather than "for the duration of the flight / pendant tout le vol" which would be "na tango tî <u>hürü</u> nî kwê".

19. Une <u>notice de sécurité</u> placée devant vous est à votre disposition.

We encourage everyone to read the <u>security leaflet</u> located in the <u>seatback pocket</u>.

Mbênî <u>mbẹ̈tï̈wängö tî̈ bata-sîrîrî̈</u>ayeke na yâ tî̈ <u>bozöbëkë</u> sô na gbelê tî̈ mo. Nzönî mo dîko nî ngâ.

This is another example of the diversity in the observation of reality. In French "notice de sécurité" puts forwards the "information" side of the document whereas in the English wording "security leaflet" it is the support of this information that is emphasized. In Sängö, "mbëtïwängö", literally "paper-advice" combines the two aspects. In the French sentence, it is not specified where the security leaflet is placed as it is mentioned in the English version. So, we chose to translate "seatback pocket" in Sängö to deliver a more precise message. The Sängö term "bozöbëkë" is made of "bozö" pocket and "bëkë" seatback.

20. Merci pour votre attention. Nous vous souhaitons un bon <u>vol</u>.

Thank you for your attention. We wish you a very pleasant <u>flight</u>.

Sîngîla sô âla mä ë sô. Nzönî <u>hürü</u>naâla.

This is a specific context in which the neologism "hürü" (flight / vol) can be used and understood. The short noun phrase "Nzönî hürü na âla" meaning "Good flight to you" makes it possible to guess and learn that "hürü" means flight, as the verb "huru" is already very common and well known. Many other couples of words in Sängö have this feature of tonal opposition between verbs and nouns derived from the same semantical and morphological roots.

2. The Resulting Terminology List

Here is the resulting terminology list from the above translation work.

French	English	Sängö
01. À bord	On board	Na yângö
02. Accoudoir	Armrest	Wotï
03. Appareil électronique	Electronic device	Fonônö tî dadä
04. Attacher (ceinture)	Fasten (belt)	Kânga (darakûba)
05. Atterrir	Land	Zunda
06. Atterrissage	Landing	Zündängö
07. Boucle	Buckle	Bïngï
08. Décoller	Take off	Löndö
09. Décollage	Take off	Löndöngö
10. Dépressuration	Decrease in cabin pressure	Pëtëpupu (tî yâ tî lapärä) atîa
11. Détacher (ceinture)	Release (belt)	Zâra (darakûba)
12. Dossier (desiège)	Seat back	Bëkë (tî kitî)
13. Dossier redressé	Seatback in an upright position	Bëkë tî kitîalütï
14. En cas d'urgence	In case of emergency	Na gbâgbûru
15. Équipage	Cabin crew	Áwakua tî yâ tî lapärä
16. Gonfler	Inflate	To pupu / Gbôto pupu
17. Gilet de sauvetage	Life jacket	Kangalïngäbatafi
18. Il est strictement interdit de fumer	It is strictly prohibited to smoke	A ke kâsâ kâsâtîtene zo anyön mânga
19. Issue de secours	Emergency exit	Yângâda tî sô kwâdaâ
20. Libérer l'oxygène	To release the oxygen flow	Tîtene tâsôkö nî asua
21. Marquage lumineux	Pathway enlightened (on the floor)	Zängö lêgë tî wâ (na sêse)
22. Masque à oxygène	Oxygen mask	Tagï tî tâsôkö
23. Notice de sécurité	Security leaflet	Mbëtïwängö tî bata-sîrîrî
24. Panneau lumineux EXIT	Illuminated EXIT sign	Zängö wâfâ EXIT
25. Poignet rouge	Red tackle	Bengbäligbô
26. Sécurité	Security	Bata-sîrîrî, bata-terê, sîrîrî
27. Siège	Seat	Kitî
28. Téléphone portable	Mobile phone	Sînga tî bozö
29. Toboggan	Emergency slide	Ngözënë
30. Vol	Flight	Hürü

3. Conclusion

As mentioned in the introduction to this paper, to elaborate this terminology, I have applied the method of a culture based approach to terminology which I have initiated and developed with my colleagues during ten years (1998–2008) in the *Laboratoire des Langues et Cultures d'Afrique Noire (LLACAN), Centre National de la Recherche Scientifique (CNRS,) Paris, France* [Diki-Kidiri et al. 2008]. Using the same method, we have elaborated reliable and sustainable terminologies in a large variety of specialized domains such as justice, administration, mathematics, agriculture, finance, elections, linguistics, computer science, etc.

This actual paper is nothing but a short sample to show how less used languages can be incapacitated for a larger use in new technical domains closed to them up to now. The next step is to completely cover the full range of the needs of announcers. They want not only security announcements but also commercial and technical messages as well. Once the work of elaborating the terminologies is completed, a very important step is still to follow. This is the training of announcers. Native speakers of a less used language are not usually comfortable when they have to use it for the first time in a specialized domain their language is not usually used for. It takes some time to seriously train them till they become fluent users of this professional variety of their language. Ultimately, a large public, in our case all passengers, who is exposed to this professional variety of the language progressively gets familiar to it and finally understands it good enough to become comfortable with it. Like this, not only the language is incapacitated, but also the professionals and finally the public at large who are ordinary speakers of that language.

References

1. Diki-Kidiri, M. et al. (2008). *Vocabulaires scientifiques dans les langues africaines. Pour une approche culturelle d ela terminologie.* Editions KARTHALA, 299 p.

2. Diki-Kidiri, M. (2014). Le vocabulaire juridique en sängö. In: Temmerman, Rita and Marc Van Campenhoudt (eds.), *Dynamics and Terminology: An interdisciplinary perspective on monolingual and multilingual culture-bound communication.* John Benjamins 2014. vi, 305 pp. (pp. 99–110).

3. Tourneux, H. (2006). *La communication technique en langues africaines.* Editions Karthala, 157 p.

4. Tourneux, H., Boubakary, A., Hadidja, K. (2011). La transmission des savoirs en Afrique. Savoirs locaux et langues locales pour l'enseignement. Editions Karthala, 304 pages et 1 CD-Rom.

Ludovit MOLNAR

President, Slovak National Commission for UNESCO;
Professor, Slovak University of Technology
(Bratislava, Slovakia)

IIT Approach to Linguistic and Cultural Diversity in Cyberspace

Summary

There are many approaches to solve real or potential problems of multilingualism in cyberspace. One of them is Informatics and Information Technology (IIT) approach. Motivation is very straightforward : IIT is "behind" cyberspace technology. IIT is also "behind" our Information for All Programme. "IIP" is one of the predecessors of IFAP. In this paper we will analyze languages used in the IIT environment, their properties as carriers of information and/or knowledge, their properties required for description of computation, their properties as communication tools in communication of human beings with computer, as well as computers within computer network. We will present the experience with intermediate language in the process of translation or interpretation in IIT. An analysis of natural languages use and processing by IIT is also done. We will show the role of language for the *availability* of information and/or knowledge as an inevitable part of information and/or knowledge *access*. We will also describe some potential next steps to make available information and/or knowledge accessible.

Introduction

Informatics and Information Technology (IIT) is "behind" cyberspace technology and also "behind" IFAP. Furthermore language and multilingualism are nothing new for the IIT environment. Using a computer requires for communication of the user with the computer and multilingualism is used in the IIT environment as well.

Language enables presentation and dissemination of information, language enables information exchange. *Presentation* and *dissemination / exchange* of information is a strong part of IT. *Communication* has been completely changed by IT.

Language as a Basic Communication Tool

In the IIT environment language is a basic communication tool for the communication of a human being with the Computer, and for the communication within computers in a computer network.

Computer also has its "mother language". Knowing this language is for communicating with the Computer. This is valid also for the communication within computers. "Mother language" of the Computer – machine code – is good for the Computer (Computer architecture), but it is "too low" for the user – a human being. As a consequence computer programming in mother language – machine code – is difficult and inefficient. To overcome this difficulty high level programming languages have been developed. They are more suitable for the user, but communication with the Computer in such a language requires an intermediate step – translator / interpreter. Actually communication with the Computer in a high level programming language is more about "what" (content of communication) than about "with whom" this communication is realized, more about programmes, their construction and corresponding computation realized by the Computer. Some representatives of high level programming languages: Machine code, Assemblers, in 1951 Grace Hopper programmes the first compiler A-0, paving the way for the higher level programming languages used today, FORTRAN (FORmula TRANslation), LISP, COBOL, ALGOL 60, BASIC, C, Pascal, Smalltalk, Prolog, Adobe and PostScript, Perl, Java, Python...

Preservation of Knowledge

Programmes in any programming language (PL) represent knowledge associated with the given PL. Similarly as in the case of natural languages, if a given PL is not used / replaced by a new PL, we *lose* all knowledge represented by programmes prepared in the given PL.

Formal, Natural and Programming Languages

Formally, language is a subset of set of all strings created from the symbols of a given alphabet. Let A be an alphabet, A^* – a set of all possible strings of symbols from A (including empty string), then language L is a subset of A^*. A programming language is a formal language with specific properties. The PL is specified by the corresponding grammar. Communication of a user with the Computer in a PL is realized by a programme and corresponding computation. It is just computation which must be equivalent at each level of the PL and finally realized through the mother language of the given computer. An hierarchy of languages can be built up through "Sets of strings" and through

"Computations". Both approaches are important in communication and also in education.

Communication with the Computer has some specificities:

- Computer is unable to communicate in a *non-mother language,*
- Communicating with computer we cannot rely on "common sense" ,
- Communication language must be *deterministic* and *unambiguous*

As a consequence to communicate with the Computer in a non-mother language we need "somebody" who knows both languages to translate or to interpret.

A number of high level programming languages and a subsequent need of translators or interpreters led to some "intermediate" languages (IL). They brought some "savings" to the translation process that can be seen from the following scheme, where SL is a source language and GL is a goal language:
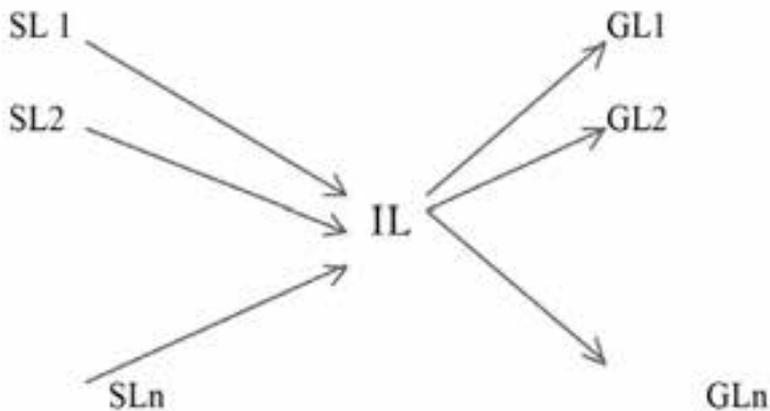
Intermediate language uses to have properties like "simpler translation *from SL*", "simpler translation *to GL*". Also operations "*on IL*" use to be simpler.

Accessibility and Availability

High level programming languages have changed the communication with the Computer, but computers themselves (IIT) changed communication between human beings. IIT not only enables different representation of information exchanged (voice, text, picture, etc.), but enriches the position of sender and receiver of information as well. You don't have to send information directly to the receiver, it can be presented in a special place instead and this way become *available* for the receiver. We can say that communication has been changed from "receiving information" to "information search". As for a language used

for information representation, the situation is the same: the receiver *needs to understand the language used for the information representation.* This concerns also natural languages.

Information search brings another notion – *accessibility.* It is natural that information is "*accessible*" only if it is "*available*". While information availability is connected with the information sender and reflects the language which he understands, information accessibility is connected with the information receiver. If he wants to use the accessed information, he needs to understand the language in which it is represented or he needs help – "someone" who can translate or to interpret it.

IIT and Linguistic and Cultural Diversity in Cyberspace

IIT can "help". It is important for the availability of information / knowledge in a given language. It is important for preservation of information / knowledge in a given language. It is important for different ways of representing information. It is important for translation or interpretation from a given to the required language.

Claudia SORIA

Researcher, National Research Council
(Pisa, Italy)

Towards a Notion of "Digital Language Diversity"

1. Introduction

How many people can use their mother tongue online extensively and without encountering any problems? What are the languages mostly used on the Internet and across digital devices? Speakers of regional and minority languages, or even speakers of a language other than the ten most used over the Internet[18], experience several problems when trying to use digital devices. For instance, it can happen that not much content is available in that language. Or the keyboard of the PC is not equipped with the characters and diacritics necessary to write in the language. Or else, it might be the case that there is too much embarrassment in showing to the world that one's written competence is not up to standard. Of course, speakers of a minority language are also speakers of a majority one, and they could use that language to access the Internet. But what are the implications of this choice, that in many cases is a forced one? What are the conditions for a language to be used online? And what is the linguistic diversity of the Internet?

2. Linguistic Diversity

According to linguists, there are between 6,000 and 7,000 spoken languages [Lewis et al. 2013], and perhaps as many sign languages. The impressive language diversity of the world is reported to concentrate in some areas more than in others: for instance, Papua New Guinea (home to 830 languages over 400,000 km^2), Indonesia (722 languages for 240 million people), Nigeria (more than 500 languages), India (22 official languages, 400 languages, more than 4000 dialects). These areas of incredible concentration of different languages are called *language hotspots*: regions having not only the highest levels of linguistic diversity, but also the highest levels of endangerment, and often the least-studied languages [Harrison 2010a].

The western world has long been biased by the myth of the Tower of Babel: linguistic diversity is a curse to fight against, and monolingualism is the cure

---

[18] *According to the Internet World Stats (www.internetworldstats.com) they are: English, Chinese, Spanish, Japanese, Portuguese, German, Arabic, French, Russian and Korean.*

for a more peaceful and harmonized world. Western people have interiorised their own monolingualism, based on the nation-state philosophical concept and political incarnation, and tend to make the monolingual regime appear the main linguistic experience of the world. At the same time, monolingualism is believed to be a guarantee of a functioning world order, and a modernizing force. Modifying this view is extremely difficult, since language lies at the heart of identity building, personal as well as national (of an individual as well as of a nation): asserting the equal value of different languages is a narcissistic wound and questions the political relationships.

The monolingual mindset stands in sharp contrast to the lived reality of most of the world, which throughout its history has been more multilingual than unilingual. Linguistic diversity, not monolingualism, is the normal, natural condition of the relationship of humankind with its surrounding environment. Language diversity is the human response to the variability of natural environment, and in places where the density of different languages is very high, such as Papua New Guinea, Central America, Africa and the Far East, it is absolutely normal for people to speak several languages. The variety of the ways in which human beings have adapted and responded to the various climates and challenges is uniquely embodied in languages. As such, it represents an important guide to understanding the interactions of humans with nature. The parallelism with biodiversity has repeatedly been made: it appears that those places with high species diversity (tropical forests in particular) tend to show high linguistic diversity, while areas low in species diversity, such as deserts and tundra, also show low linguistic diversity [Loh and Harmon 2014; Nettle and Romaine 2000; Loh and Harmon 2005]. But there are more similarities between linguistic and biological diversity besides their distribution: both are facing an extinction crisis, and both crises are consequences of similar processes. Exactly as it happens for biodiversity, language diversity is severely endangered, in some places more than in others [Loh and Harmon 2014; Harmon and Loh 2010]. According to Sutherland [2003], the loss of languages goes at a faster pace than the loss of species. The reasons behind the loss of linguistic diversity are mostly concerned with social or economic issues (commerce, migration, globalization of trade and media, but also unfavourable national policies and the prestige associated with one or more dominant languages); more rarely they are associated with natural phenomena such as a population's extinction.

2.1. Protection of Linguistic Diversity

Exactly like biodiversity, linguistic diversity is a heritage to be preserved by all means, not a problem to be eradicated. Strangely enough, people – at least western people – tend to recognize the value of biodiversity much more

112

than they do for linguistic diversity. They may be keen on protecting whales and wolves from extinction, but could not care less if the language of their grandparents will disappear by the next generation. If we believe that language diversity is a value, we need to support it as our collective responsibility towards humankind. Languages are the living archive of human experience: a monument of the peculiarly human way of forming societies, communicating, and transmitting experience.

David K. Harrison, a linguist and advocate of linguistic diversity, expresses this view in a very powerful way: "What hubris allows us, cocooned comfortably in our cyber-world, to think that we have nothing to learn from people who a generation ago were hunter-gatherers? What they know – which we've forgotten or never knew – may some day save us. We hear their voices, now muted, sharing knowledge in 7,000 different ways of speaking. Let's listen while we still can." [Harrison, 2010].

If we think that remote languages and cultures cannot teach us anything, simply because they have never seen a mobile, we are just wrong. Humans have spent centuries in close interaction with often extremely harsh and demanding environments, and their languages encode knowledge that might turn useful, someday or another: knowledge of surviving techniques, of plants, animals, and crops, preparation and uses of medicinal food, traditional methods of farming, fishing and hunting, not to mention traditional methods of land use and resource management. We cannot afford to lose this enormous wealth of knowledge that was accumulated over the centuries. *Let's listen while we still can.*

2.2. Sustainment of Linguistic Diversity in the Digital World

In order to establish a sustainable policy for safeguarding and promoting linguistic diversity, the digital world cannot be ignored any longer. As Mark Turin aptly says, "in our digital age, the keyboard, screen and web will play a decisive role in shaping the future linguistic diversity of our species" [Turin 2013]. Languages are living entities that need to be used on a daily basis by humans in order to survive.

With so much of our lives happening on the Internet and through digital devices, the digital space represents a context that cannot be ignored. Speakers of major languages can access apparently unlimited amounts of Web content, easily perform searches, interact, communicate through social media and voice-based applications. They can enjoy interactive e-books, have fun with word games for mobiles, engage in multi-player videogames, or take advantage from innovative language learning facilities for other widely spoken languages.

On the other hand, speakers of minority languages cannot benefit of any similar facility. So called "smaller" languages do not enjoy the same range of opportunities. Welsh speakers were denied the publication of e-books in Welsh over Amazon's Kindle platform, because of lack of available Welsh electronic dictionaries. There is no Wikipedia for Mansi; speakers of Saami or Tongva have no localized interface for Facebook, and there is no Google translation for Sardinian, or Igbo, or Frisian. This *inequality of digital opportunities* further discriminates minority languages, by relegating them once more to the realm of family communication and restricted topics. Minority languages, instead, need to get access to all contexts of life to be perceived as vibrant and fully apt languages. Presence of a language on the Internet is of paramount importance for the impact it has on its speakers, especially the young generation. We must ensure, therefore, that the range of usage opportunities for all languages is increased and enlarged. Multilingualism cannot be truly and effectively enforced if all languages are not put in the conditions to act digitally. Empowering all languages, regional and minority ones in particular, with instruments that put them on a par with more widely spoken languages, is a matter of equal digital opportunities for the speakers of those languages. *Digital Language Diversity needs to be sustained.*

3. Is the Internet Linguistically Diverse?

According to a recent survey (LTInnovate), in 2012 digital content has grown to 2.837 zettabytes, up almost 50% from 2011, on its way to 8.5 ZB by 2015. The community of social network users in Western Europe was set to reach 174.2 million people in 2013, which is about 62% of Internet users. A massive of 800 million people are Facebook users, of which 170 millions are from highly linguistically diverse countries such as Brazil, India, Indonesia, and Mexico. The number of Twitter's active users is estimated around 200 millions. LinkedIn has 115 million users, and Google+ as many as 180 millions[19].

These numbers, as imperfect as they may be, give a flair of the depth and breadth of the Internet, but what can we say about its linguistic diversity? How the enormity of Internet users behave, from a linguistic point of view? Which languages do they use? In other words, does the Internet reflect the linguistic diversity of the planet?

A study by W3Techs[20] shows that at the time of writing of this article, 55.9% of all content online is in English. Aside from English, Spanish and Portuguese,

---

[19] *Source: Language Connect: www.languageconnect.net.*

[20] *http://w3techs.com/technologies/overview/content_language/all.*

only five other EU languages (German, French, Italian, Polish and Dutch), out of 60 or more spoken in the Union, are published on more than 1% of the top million sites[21].

With reference to domain names, a majority of domains (78%) are registered in Europe or North America: a finding that reinforces the dominance of those two regions in terms of Internet content production. Asia, in contrast, is home to 13% of the world's domains while Latin America (4%), Oceania (3%), and the Middle East and Africa combined have even smaller shares of the world's websites (2%). Globally, there are about 10 Internet users for every registered domain. The United States is home to almost a third of all registered domains, and has about one website for every three Internet users.

From the Wikipedia point of view, Wikipedia articles in 44 language versions of the encyclopaedia are highly unevenly distributed. Slightly more than half of the global total of 3,336,473 articles are about places, events and people roughly concentrated in the European area, occupying only about 2.5% of the world's land area: the majority of content produced in Wikipedia is about a relatively small part of our planet.

The Internet is not as linguistically diverse. English is still the language most used over the Internet, the one for which more content is produced, and also the privileged tongue of the majority of its users.

3.1. The (Slowly) Growing Linguistic Diversity of the Internet

The preminence of English, however, is being rapidly eroded: according to a 2012 survey[22], which used the users' origin as proxy of the languages used, English has diminished from 39% in 2009 to 27% in 2011.

There are 22 domain names across the world, with 100 more expected to go live soon. There are more than 160 million websites globally, but about 111 million of them end with *.com*. Of the 2 billion Internet users, more than 70 per cent are not native English speakers. Another 2 billions are expected to go online by 2016, almost all of whom will not count English as their first language.

As for the language of social media, the so-called "informal Internet" remains a safe haven for minority languages, thus confirming the intuition expressed by Daniel Prado in 2008 [Prado 2011]. The Indigenous Tweets site[23] tracks tweets in 149 different languages, over a total of 61,828 accounts and more than 12

---

[21] *Source: LTInnovate.*

[22] *Source: Smartling, www.smartling.com.*

[23] *http://indigenoustweets.com/.*

million tweeters. Of these, 33 languages have a single tweeter only, somehow a digital counterpart of so-called "last-speakers".

The other massively spread social site, Facebook, has about 83 "official" translations, but the personal profiles, groups and pages show an immensely higher linguistic diversity.

The Internet is not as linguistically diverse as the "real world". There is a huge disproportion between the languages actually spoken and those represented on the Internet. Scannell [Scannell 2013b], reporting about the Crúbadán work in progress, has traced back as many as 1,510 languages over the Internet. Should this figure be increased, as Scannell himself suggests, to even 1,800 languages, it would mean that a mere 26% of the world's languages are represented over the Internet.

It is plain that the Internet will never be able to perfectly mirror the actual world's linguistic diversity, either for connectivity reasons or for the simple fact that only a few hundred languages have a writing system (between 5% and 10% of world languages, according to sources). Also, languages using Latin characters have been favored over others, simply because the Internet was at the beginning a tool created to suit the English language. Not all languages have the same possibilities of getting represented over digital tools. It is important to reflect on the implications of this digital divide. The increase in availability of smartphones and digital connectivity will determine an increase in the demand of content and services offered in a multitude of languages. And indeed, the Internet is responding, and slowly growing linguistically diverse. In order to account for a growing linguistically diverse market, Google amplified its language offer, from 43 languages in 1998 to 80 in 2014. Facebook currently supports 83 languages, and Twitter 33. The question is: *is the pace as fast as necessary?*

4. Digital Language Diversity

The concept of *Digital Language Diversity* is an extension of the concept of Language Diversity to the digital realm. As such, it aims to capture the amount of languages over a given digital population, tools, and applications.

Digital Language Diversity is important under many respects. The first is a matter of linguistic rights, and equal digital opportunities for all languages and all citizens.

The second is connected to heritage preservation: digital tools allow the documentation of languages, and hence the preservation of their cultures, in a way that was not precedented (much faster, much safer). However, preserving

a language is like putting a precious tool in a museum: it might be preserved, polished and restored, it might be admired by many visitors, but it will never be used again. Languages need to be used to be vital, and a language that is not used with digital tools is no longer a fully apt language.

Therefore, Digital Language Diversity is important also for identity reasons, and for allowing people to take pride in their language. There are also economic implications: the more services are offered in more languages, the more people and more consumers will be reachable as the digital market expands.

4.1. Digital Language Diversity and Language Under-Representation

A low level of Digital Language Diversity means that many languages are *under-represented*.

The concept of under-representation basically applies to any language that suffers from a chronic lack of resources, be they human, financial and time resources or linguistic resources (language data and language technology).

On the whole, we can distinguish between two main aspects of under-representation: a) as content and b) as uses. Content under-representation means that no or very little content is available in a given language; uses' under-representation means that although there is some digital content in a given language, it is a merely static one: it is not possible to do anything with it – there is no localized interface, there are no services available in the language, and a user cannot really interact using the language over digital devices. He can only access some web pages. A typical case is when there is a Wikipedia in a given language, but not localized interfaces of most popular applications and programmes: in order to access the Internet and take profit of the services available on it, a user must switch to another language.

It will be no surprise, therefore, that the majority of languages are under-represented according to this definition.

4.2. "Digital Diglossia" and "Digital Exctinction"

A language that is under-represented is a language that has less contexts where it can be used, and less opportunities to be used than other languages have.

Less digitally represented languages are under a serious risk of being marginalized, and eventually *dialectalized* over the years.

According to Carlos Leáñez (cited in [Prado 2011]), the less valuable a language is [in the eyes of its speakers], the less it is used, and the less it is used, the more it loses value". Shrinking contexts of uses can have a devastating effect,

eventually leading to the abandonment of a language in favour of another, better supported one. Should this happen, the consequences for a language profile would be dramatic: any language that cannot be used over digital contexts will engage in a "digital diglossia" relationship with another, better supported language.

Not only those languages that struggle to get access to the digital world, but even languages that are digitally represented at the mount are at risk. Less and less digital contexts of use is what can bring languages to *digital extinction*. It is common to associate the concept of extinction with very exotic languages, or those spoken by a restricted minority. However, the concept of "digital extinction" describes a condition that could prove true for many languages, even those far from being endangered outside the digital world. This condition holds whenever a language is used less and less over the Internet because of lack of Language Technology support: then the range of contexts where it is used dramatically collapses and gradually brings the language to disappear from the digital space.

Where there is no favorable environment for a language over digital tools, its use over the Internet and through digital devices becomes cumbersome, communication is difficult, and usability of the language is dramatically affected. By pushing the naturalistic metaphor further, we can think of a "digitally hostile environment": one where it is not possible to type, make searches, have translations, hold a conversation over digital devices. In such a context, a language easily goes extinct.

The concept of digital extinction was first introduced by a research carried out by the META-NET Network of Excellence[24], culminated in the publications of 30 "Language White Papers" [Rehm and Uszkoreit 2012], one for each official EU language. This research, which is freely accessible and downloadable from the META-NET website, reports about the current and future state of the languages with respect to their technological support, and has had a strong impact in the press and helped structure the current framework of European funding.

The study includes a comparison of the support all languages receive in four application areas: machine translation, speech processing, text analytics and language resources. The differences in technology support between the various languages and areas are dramatic and alarming: Language Technology support varies considerably from one language community to another. In

---

[24] *www.meta-net.eu.*

the four areas, English is ahead of the other languages but even support for English is far from being perfect. While there are good quality software and resources available for a few larger languages and application areas, others, usually smaller languages, have substantial gaps. Many languages lack basic technologies for text analytics and essential resources. Others have basic resources but semantic methods are still far away. A recently update of the study [Rehm et al. 2014] demonstrates, drastically, that *the real number of digitally endangered languages is, in fact, significantly larger*.

5. Preventive measures

How can Digital Language Diversity be fostered and Digital Extinction prevented?

Using the words of John Hobson (quoted by Kevin Scannell [2013]), "The Internet and digital world cannot save us. They cannot save Indigenous languages. Of course these things have benefits but they are not the Messiah. We don't need another website or DVD or multi-media application, these are short term, quick fix solutions. What we really need is sustainable initiatives, to create opportunities for Indigenous language users to communicate with each other in their native tongue. To get people speaking again."

The META-NET study described above clearly shows that, in our long term plans, we should focus even more on fostering technology development for smaller and/or less-resourced languages and also on language preservation through digital means. Research and technology transfer between the languages along with increased collaboration across languages must receive more attention.

Although the destiny of a language is primarily determined by its mother-tongue speakers and its broader cultural context, the technological development of an under-resourced language affords the language the strategic opportunity to have the same "digital dignity", "digital identity" and "digital longevity" as large, well-developed languages in the Web.

If we want to save and preserve language diversity, and especially minority and regional languages, we must necessarily let these lesser-used languages have access to the tools and resources of the same technological level as those of "bigger" languages. The moment is now: if we don't act quickly and effectively now, if carefully planned and focused intervention is not immediately carried out, it might be too late.

5.1. The Opportunities and Challenges of Language Technologies

Language-based applications are at the very core of Digital Language Diversity. The market of such applications and services is increasing day by day, and the new digital tools are doing wonders for endangered languages, by offering a way back from the brink for a many languages that seemed doomed just a few years ago. There are several examples of how new media and digital technologies are helping the salvation of moribund or endangered languages: North American tribes use social media to re-engage their young, and there is an iPhone app to teach new students the pronunciation of Tuvan words (an indigenous tongue spoken by nomadic peoples in Siberia and Mongolia). An app for Tusaalanga Inuktitut is being developed as a resource for learning several Inuktitut dialects.

The essence of language-based applications is Language Technology (LT), i.e. data and software that allow the automatic processing of natural language, such as spelling and grammar checkers, electronic dictionaries, localized interfaces, as well as search engines, automatic speech recognition and synthesis, language translators or information extraction tools. LT can offer an enormous help to minority languages, for example by offering a writing aid, by helping spread a language among children by means of apps and e-books, by building electronic dictionaries that are far more easily usable than traditional ones. If we accept that modern Information and Communication Technology (ICT) is indeed an opportunity for small languages, we must recognize that on the other hand it constitutes a big challenge, as it requires fast development of high quality LT to keep up the pace of technological development.

In other words, ICT will indeed help minority languages in their gaining a space over the digital place, but only on the condition that good and effective LT is developed and integrated into ICTs. If a language is not adequately supported by language technologies, its use over the Internet and through digital devices becomes cumbersome, communication is difficult, and usability dramatically affected. Development of Language Technology thus becomes an important – critical – part of language preservation and revitalization.

5.2. Language Digital Survival Basic Kit

It is by no means simple for a minority language to get engaged in the digital world. Small languages need to be given the voice, in technological terms. The challenges – ranging from digital divide and connectivity access, problems in terms of scripts and their digital encoding, lack of terminology, etc. to availability and development of language technologies – are daunting. However, going digital is not impossible for languages, as long as some minimal conditions are

met. A very basic kit ensuring a minimal degree of "digital survival capacity" for any language includes at least the following (in increasing order of necessity):

- ensured connectivity ;
- a sufficiently developed and adopted standardized encoding;
- a developed terminology;
- a standardized orthography;
- localized interfaces;
- basic language resources are available, at least including a corpus, spell checker, and lexicon

The Language Digital Survival Basic Kit might be considered as a remodelling of the notion of BLaRK (Basic Language Resource Kit, [Krauwer 2003]), i.e. the minimal set of language resources that is necessary to do any precompetitive research and education. The BLaRK lists, for a given language and for several different language technologies applications, the data and software modules that represent a prerequisite for those technologies. Although, in principle, this is a language-independent concept, its instantiation heavily depends on the specific requirements of individual languages. We can think of a BLaRK as a LRT "checklist": with this list in one hand and an updated catalogue of the available resources in the other, it becomes possible to effectively make a development plan, prioritized according to the different needs of different languages, for endowing less resourced languages with a minimal "digital survival kit".

6. An Index of Digital Language Diversity

Protection of cultural and language diversity imposes that, in a world dominated by ICT, all communities, all languages, all cultures be first class citizens. The challenges for attaining this status can be haunting for many languages; yet, there is no way back from entering the digital realm for any language that truly aspires at being a vital one.

In order to assess Digital Language Diversity, either on a global or local scale, we need reliable indicators that allow to determine the status of languages vis-à-vis their digital representation and vitality.

If we closely follow the work recently being done in conservationist biology [Loh and Harmon 2014], in order to categorise the conservation status of a digital language we need the following indicators: i) a digital language's

population size; ii) its rate of reduction or expansion; iii) its range size and rate of decline or fragmentation; and iv) existing and future threats.

We have seen how the META-NET Language White Papers describe a methodology for assessing the digital extinction risk of a language (point iv) above). The measurement of the size of digital languages' "population" was the explicit objective of the "Language Observatory"[25] centre set up at Nagaoka University, Japan [Mikami and Nakahira 2011], which used languages, scripts and encoding of web pages as proxies for language diversity.

Exactly like conservationist biologists are interested in alive plants and species, a measurement of Digital Language Diversity needs to focus on vital languages, not dead ones. Yet the Web can host heritage languages, such as corpora of Ancient Greek or Old Saxon poems. Reliable indicators of the vitality of a digital language, i.e. of the extent to which a language is digitally prosperous by increasing its presence online, are still missing. Kornai's "Digital language death" [Kornai 2013] represents the first attempt at devising reliable indicators of Digital Vitality (points ii) and iii) above) by bringing the traditional methods of language vitality assessment to the digital realm. In doing so, he correctly identifies active digital uses of a language as a crucial factor in determining its Digital Vitality, and therefore suggests to complement the indicators of digital presence of a language (i.e. number of web pages in a given language) with other proxies for digital language use, such as the existence of an active Wikipedia community in the language.

As a preliminary extension to Kornai's and others' previous work, we propose the following list of indicators of any language's healthy Digital Language Vitality:

1. big size of digital population (in terms of Facebook/Twitter accounts and considering population between 10 and 70 years of age);

2. use by global brands (e.g. Google, Microsoft, Apple, etc.);

3. strong Internet penetration;

4. big Internet content (in terms of number of websites and number of websites per speaker);

5. the most visited websites of the country and of the world have a localized version;

6. availability of a Wikipedia;

---

[25] *http://gii2.nagaokaut.ac.jp/gii/lopdiary.php?blogid=8&catid=109*.

7. social media have a localized interface;

8. the language is used on blogs, twitters and other social communication tools (e.g. email, chat, forums, discussion lists);

9. the language has a dedicated Internet domain;

10. the main operating systems have a localized version;

11. there are language apps available in the language;

12. there are machine translation tools for that language.

Substantial work is necessary in order to work out these indicators in detail, especially in order to associate appropriate proxies for indicators such as 1), 3) and 8), as well as to develop reliable methodologies to measure 4) and 8).

However, the idea of measuring and assessing the linguistic diversity of the Web has been around for quite a long time now, and we believe that the time has come to converge towards concrete actions, especially since affordable and open methodologies have appeared in the meantime, most notably the Cr b dan project [Scannell 2013b] and Kornai's work [Kornai 2013]. We would favor the establishment of a collective effort, possibly under the UNESCO's aegis, to advance work in this area towards the development of such an Index of Digital Language Diversity.

7. Conclusions

Widening Digital Language Diversity is desirable and possible, as there is no limitation, in principle, to the number of languages accessing the Internet and content provided in those languages. Even if Digital Language Diversity will never be able to mirror the world's linguistic diversity, we can and should aim at least at a partial reflection of it. International and national policy makers should support and foster the digital presence of minority languages, in particular those more at risk of digital extinction. The range of technical and political challenges involved is very vast, and must be addressed at once in order to endow languages with the minimal necessary instruments in order to access the Internet and start producing content. The development of reliable indicators of Digital Language Diversity is also desirable and we argue that such an initiative should be collectively and collaboratively pursued, possibly under the aegis of UNESCO. These indicators could be used to build an Index of Digital Language Diversity, to be used as a monitoring tool to assess digital language diversity in a certain area and highlight areas where intervention is needed (for instance, by singling out where effort should be channelled and funding directed).

Although the destiny of a language is primarily determined by its mother-tongue speakers and its broader cultural context, a Digital Language Strategy could help directing the technological development of an under-resourced language, thus affording the language the strategic opportunity to have the same "digital dignity", "digital identity" and "digital longevity" as large, well-developed languages in the Web.

References

1. Harmon, D., Loh, J. (2010). The index of linguistic diversity: A new quantitative measure of trends in the status of the world's languages. *Language Documentation and Conservation*, 4, pp. 97-151.

2. Harrison, D. K. (2010). *The Last Speakers*. National Geographic, Washington, DC.

3. Harrison, D. K. (2010). *The tragedy of dying languages*. BBC News. http://news.bbc.co.uk/2/hi/8500108.stm.

4. Kornai, A. (2013). Digital language death. *PLoS ONE*, 8(10).

5. Krauwer, S. (2003). The Basic Language Resource Kit (BLaRK) as the First Milestone for the Language Resources Roadmap. *Proceedings of the 2003 International Workshop Speech and Computer* (SPECOM 2003), pp. 8-15.

6. Lewis M. P., Simons G. F., Fennig C. D. (eds.) (2013). *Ethnologue: Languages of the World*. 17th edition. SIL International, Dallas, Texas, USA.

7. Loh, J., Harmon, D. (2005). A global index of biocultural diversity. Ecological Indicators, 5, pp. 231-241.

8. Loh, J., Harmon D. (2014). *Biocultural diversity: threatened species, endangered languages*. WWF, Netherlands.

9. LT-Innovate.eu. (2013). *Status and potential of the European language technology markets*. LT-Innovate Report, March 2013.

10. Mikami Y., Nakahira K. T. (2011). Measuring linguistic diversity on the Internet. In: E. Kuzmin and E. Plys (eds.), *Linguistic and Cultural Diversity in Cyberspace. Proceedings of the International Conference*. Interregional Library Cooperation Centre, Moscow, pp. 136-144.

11. Nettle, D., Romaine, S. (2000). *Vanishing voices: the extinction of the world's languages*. Oxford University Press, Oxford.

12. Prado, D. (2011). Languages and cyberspace: Analysis of the general context and the importance of multilingualism in cyberspace. In: E. Kuzmin and E. Plys (eds.), *Linguistic and Cultural Diversity in Cyberspace. Proceedings of the International Conference.* Interregional Library Cooperation Centre, Moscow, pp. 72-82.

13. Rehm, G., Uszkoreit, H. (eds.) (2012). *META-NET White Paper Series: Europe's Languages in the Digital Age*, Springer, Heidelberg, New York, Dordrecht, London.

14. Rehm, G., Uszkoreit, H., Dagan, I., Goetcherian, V., Dogan, M. U., Mermer, C., Váradi, T., Kirchmeier-Andersen, S., Stickel, G., Prys Jones, M., Oeter, S., Gramstad, S. (2014). An Update and Extension of the META-NET Study "Europe's Languages in the Digital Age". In: *Proceedings of the Workshop on Collaboration and Computing for Under-Resourced Languages in the Linked OpenData Era* (CCURL 2014), Reykjavik, Iceland, May 2014.

15. Scannell, K. (2013). Endangered languages and social media. Presentation at the *Workshop at INNET Summer School on Technological Approaches to the Documentation of Lesser-Used Languages*, September 2013.

16. Scannell, K. (2013). How many languages are on the Web? The Crúbadán project10+ years on. Invited talk at the *Workshop on Corpus-based Quantitative Typology (CoQuaT 2013)*, August 2013.

17. Sutherland, W. J. (2003). Parallel extinction risk and global distribution of languages and species. *Nature*, (423), pp. 276-279.

18. Turin, M. (2013). *Globalization helps preserve endangered languages*. Yale-Global Online, 2013. http://yaleglobal.yale.edu/content/ globalizationhelps-preserve-endangered-languages.

Anna FENYVESI

Associate Professor of English Linguistics,
Director, Institute of English and American Studies, University of Szeged
(Szeged, Hungary)

Multilingualism and Minority Language Use in the Digital Sphere: The Digital Use of Language as a New Domain of Language Use

Introduction

In the past quarter-century, digital things have changed numerous aspects of our lives – and language use is no exception. With the use of email, the World Wide Web, mobile technologies, and digitally mediated ways of communication, a new domain of language use has entered the lives of most of us – namely, what I call "digital language use" below. A lot of it is oral, mediated by mobile phones and voice-over-IP (like Skype, for instance), but a great part of it is written and involves both reading and writing (such as emailing, texting, instant messaging, blogging, etc.). In fact, it is estimated that using these ways of digital communication, we read and write today more than before their advent [Baron 2008: 183]. This makes especially written forms of digitally mediated communication a highly important new aspect of language use that should be the focus of concern for sociolinguists, educators working in bilingual education, and, indeed, all professionals working with bilingual minority language communities, be they editors or writers of digitally present newspapers, computational linguists and computer scientists working on computational language tools, or social scientists studying the role of language in various aspects of community life.

"Domains of Language Use"

The concept of the domain of language use has been widely used in the study of bi- and multilingual communities ever since Fishman [1972: 441] adopted Schmidt-Rohr's [1932] idea of "elements of dominance configurations" as a theoretical concept defined "in terms of institutional contexts and their congruent behavioral co-occurrences. They attempt to summate the major clusters of interaction that occur in clusters of multilingual settings and involve clusters of interlocutors". The five domains originally differentiated by Fishman are family, friendship, religion, education, and employment. This set

of domains has curiously remained unchallenged ever since and is widely used by sociolinguists to this day to describe patterns of language use by bilinguals in their respective languages, indicative of language maintenance or language shift when compared intergenerationally. When linguists investigating the bilingualism of speakers discuss other domains in their work, they usually do so without explicitly stating that they have expanded their number (cf. some recent examples of domains, in Grosjean [2010: 29], "parents, children, siblings, distant relatives, work, sports, religion, school, shopping, friends, going out, hobbies, and so on"; in Bever [2011]: "public domain"; and in Öpengin [2012: 160], "economy").

Digital Language Use

But whatever the total number and range of domains to be differentiated, it seems inevitable that the "digital domain", i.e. the use of language in digitally mediated communication, should be regarded and recognized as a separate domain of language use for a number of reasons.

Digital language use has become a prominently important aspect of language use: it encompasses various forms of both formal and informal communication (rather than just the latter), it includes genuinely new functions of language use (cf. blogging), and, as scarce results already indicate, it can present patterns of language use which are markedly different from all other (traditional) aspects of language use by bilinguals. For instance, Huber [2013] has shown that while first-generation Canadian Hungarians use Hungarian more in the traditional domains of family, friendship, and religion than do their second-generation children, the latter far outperform their parents in the use of Hungarian in the digital domain (in emailing and using it on the Internet) – demonstrating that digital language use can indeed become an important factor of language maintenance for the young, "digital native" generation. Basharina [2013] has also shown that the digital domain of language use can present a space for users of the minority language Sakha in Yakutia, Russia, where new forms of old genres as well as new genres of storytelling can contribute to the strengthening of the minority language user community and its cultural and linguistic adaptation, modernization, and vitality.

In a paper meticulously supported by ample empirical data and mathematical calculations, Kornai [2013] has argued that digital language death will be the fate of a great number of the languages existing today – primarily those that exist as minority languages only – unless their speakers (and the professionals supporting them) succeed in meeting some all important criteria like having a community of digitally literate users and a Wikipedia in the language.

Support for Digital Language Use in Minority Languages

The possibility of use of minority languages in the digital domain is, of course, dependent on a number of factors which range from the technical (the existence and availability of hardware and software), through the educational (literacy in the traditional sense and in digital matters) and personal (the presence of digitally competent language users interested in using the language in this domain) to those of prestige (whether language users regard their minority language as "worthy" of the effort of using it digitally). All of these aspects present arenas where minority language users can be supported in their language use by members of their own community such as language activists and by outsider professionals – computer scientists and computational linguists working on language tools for minority languages, educators, linguists, etc. The work of the Norwegian Giellatekno company is a case in point: its computer scientists have been developing and making available a wide range of language learning tools, bilingual dictionaries, morphological and syntactic analyzers, and games for Saami and other endangered minority Finno-Ugric languages (cf. http://giellatekno.uit.no/).

An Example: The FinUgRevita Project

The "Computational tools for the revitalization of endangered Finno-Ugric minority languages, FinUgRevita" project was created in 2013 with the aim to provide computational language tools for endangered indigenous Finno-Ugric languages such as Udmurt and Mansi in Russia and to assist the speakers of these languages in using the indigenous languages in the digital domain (http://www.ieas-szeged.hu/finugrevita/).

The project involves two teams – one of the University of Helsinki, Finland (Principal Investigator: Roman Yangarber), the other of the University of Szeged, Hungary (Principal Investigator: Anna Fenyvesi) – comprising Finno-Ugrist linguists, computational linguists, and sociolinguists, and is funded for the period of September 1, 2013 to August 31, 2017 by the Academy of Finland (AKA) and the Hungarian National Research Fund (OTKA).

The two languages the project focuses on so far, Udmurt and Mansi, are both endangered, according to the UNESCO's classification of endangered languages [UNESCO 2010] although to a different extent. Udmurt is a "somewhat" endangered language, with almost 60% (or about 300,000) of the 552,000 ethnic Udmurts speaking the language (cf. the figures of the 2010 Russian census), spoken in the Udmurt Republic, or Udmurtia, west of the Ural Mountains. Even though it has official status in Udmurtia, it has limited power and rights in the public sphere and is used mostly in the family domain and among friends.

In Udmurtia it is present in the media, education and culture as well as has an Internet presence (e.g. it is one of the three Finno-Ugric minority languages that VKontakte, "the Russian Facebook" social networking site can be used in). Mansi is a severely endangered language with less than 1,000 speakers (among the 12 thousand strong ethnic Mansi population), spoken in the Khanti-Mansi Autonomous Okrug (informally known as Yugra) in western Siberia, east of the Urals. It has no official status whatsoever even in the Okrug, and although it has some minor presence in the media, education and culture of Yugra, it is used primarily in the family and friendship domains. Perhaps surprisingly for such a small language, it does have an Internet presence: the bi-weekly newspaper Luima Seripos is also published online.

Sociolinguistically, the speaker communities of both languages have been undergoing language shift, that is, the expanse of the majority language, Russian, at the expense of the minority language in the speakers' lives, ever since their ancestors came to be under Russian domination in the 16th and 17th centuries, also experiencing forceful assimilation and Russification in Soviet times [Bakró-Nagy, forthcoming]. The discovery of oil and gas in the 1970s in the regions where Mansi and Udmurt are spoken also led to the in-migration of workforce from outside, making the regions multilingual and the Mansi a minority even in their own district). For instance, the number of people professing to be of Udmurt ethnicity decreased from 640,000 in 2002 to 552,000 in 2010, while the proportion of speakers fell from 67% to 59% during the same time. And while the number of those declaring Mansi ethnicity increased in the same period, from 11,500 in 2002 to 12,300 in 2010, the proportion of speakers fell from 23% to just 7.65%.

The main aims of the project are the development of open source, freely accessible computational language tools: electronic dictionaries, morphological and syntactic analyzers, language games, as well as learning tools. Computational linguistic work on these tools has started and is in progress.

In addition to the computational linguistic work, two online surveys have been undertaken as a part of the project. One survey, launched in June 2014, aims to study the use of Giellatekno's computational language tools for Saami, with the goal of analyzing users' feedback regarding their use of and satisfaction with these tools, both for the sake of the developer company and their continuing improvement of the tools and for the FinUgRevita project being able to benefit from the experiences of the user community regarding tools similar to our future tools.

The other survey which the researchers involved in the project are preparing at the time of the writing of the present paper, August 2014, and are planning to launch in the fall of the same year is a sociolinguistic survey aimed at mapping out the digital language use of Udmurt and Mansi speakers. Specifically, through the survey sociolinguistic and language use information will be collected from speakers of Udmurt and of Mansi about what language(s) they use in various forms of digitally mediated communication, i.e. using mobile phones, emailing, surfing, chatting, blogging, commenting, using social media, producing Internet content etc. With detailed information about when speakers use the minority language (Udmurt/Mansi), the majority language (Russian), and/or other languages (English, or other minority languages spoken in Russia), it is hoped that the project's investigators will gain an invaluable insight into users' habits of language use, needs of computation language tools in minority languages, and, in general, a better understanding of language use patterns of speakers of endangered indigenous languages in the digital domain.

Conclusion

The digital domain, as I have argued above, has become an all important domain of language use by bi- and multilinguals, especially from the perspective of minority languages. Their support is essential if they are to be "digital survivors" (in terms of Kornai [2013]), although the most important prerequisite of such survival is, probably, the determination on the part of the speakers of the language themselves to save them from language shift and/or digital death – something that no outsider professional can achieve, however determined and skilled they may be.

References

1. Bakró-Nagy, M. (forthcoming). Uralic languages. *Revue Belge de Philologie et d'Histoire*.

2. Baron, N. S. (2008). *Always on: Language in an online and mobile world*. Oxford: Oxford University Press.

3. Basharina, O. (2013). *The extent of the Sakha language on the Internet: Digital genres and language revitalization*. Paper presented at the MinorEuRus conference, Helsinki, Dec. 15-17, 2013.

4. Bever, O. (2011). Multilingualism and language policy in post-Soviet Ukraine: English, Ukrainian and Russian in linguistic landscapes. http:// www.irex.org/sites/default/files/BeverSTG2011%20Final%20report.pdf.

5. Fishman, J. A. (1972). Domains and the relationship between micro- and macrosociolinguistics. In: Gumperz, J. J., and Dell Hymes (eds.), *Directions in sociolinguistics: The ethnography of communication*. Oxford: Blackwell, 435–453.

6. Grosjean, F. (2010). *Bilingual life and reality*. Cambridge, MA: Harvard University Press.

7. Huber, M. (2013). *Intergenerational transmission of Hungarian as a heritage language in Canada: The macrosociolinguistics of the Hungarian community in Hamilton, Ontario*. Szeged: University of Szeged, BA thesis.

8. Kornai, A. (2013). Digital language death. *PLoS ONE*, 8, e77056. doi:10.1371/journal.pone.0077056. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0077056

9. Öpengin, E. (2012). Sociolinguistic situation of Kurdish in Turkey: Sociopolitical factors and language use. *International Journal of the Sociology of Language*, 217: 151–180.

10. Russian census 2010. http://www.perepis-2010.ru/results_of_the_census.

11. Schmidt-Rohr, G. (1932). *Die Sprache als Bildnerin der Völker*. Munich.

12. UNESCO (2010). *Atlas of the world's languages in danger*. http://www.unesco.org/new/en/culture/themes/endangered-languages/atlas-of-languages-in-danger/.

Andras KORNAI

Senior Scientific Advisor, Computer and Automation Research Institute,
Hungarian Academy of Sciences
(Budapest, Hungary)

A New Method of Language Vitality Assessment

1. Background

In Kornai [2013] we demonstrated that over 95% of the world's languages are digitally still. This means there is a small pool of roughly 400 languages, many spoken in Russia and the FSU [Comrie 1981], from which a final set of digital survivors, perhaps some 200 languages, will emerge. Since at this point the digital ascent of no more than a few dozen languages is assured, we need a more detailed assessment than the simple four-way classification put forth in Kornai [2013] which distinguished only Thriving and Vital languages (neither Heritage nor Still languages can survive in the obvious sense of being actively used in communication). Figure 1 at the end of the paper, based on the data given in Table 1, shows this distribution for languages of the FSU, with the Thriving language (star) in the top right being Russian, and the Vital languages (circles) largely corresponding to the main languages of former republics. Squares are Heritage languages such as Old Church Slavonic, and smaller arrows corresponding to the remaining languages are either for Borderline (rightward pointing arrow) meaning that the current statistical method is incapable of fully resolving their status or for Still (down arrow), the majority of languages in the FSU.

2. Discussion

As the digital future of Thriving languages is assured, we use the lessons learned from the digital development of these to devise both a more detailed assessment of the digital potential of Vital languages and a strategy of maximizing the number of languages that make it across the digital divide. For the assessment we propose a simple log-linear formula that derives a single number D (digital vitality index) as a weighted sum of well-understood components such as the EGIDS ranking, (log) number of L1 speakers, (log) size of wikipedia, adjusted for quality, (log) crawl size, the existence of FLOSS spellcheckers, etc.

Some of the key factors, such as the number of speakers, represent long-range trends that are outside the immediate control of speakers. Others, such as EGIDS ratings, are set by expert judgment and no doubt carry some slight subjective element. From our perspective these are still objective, in that SIL experts also focus on long-range trends, such as literacy or official use,

that can be influenced only indirectly by the computational linguists primary responsible for digital vitality. These factors, which tend to be common for vitality in the traditional and the digital sense, are in sharp contrast to another group of factors we will call *volitional*. Whether there is a wikipedia, a blog, or a twitter community in a language depends on two factors: the availability of tools (entirely in the hands of software engineers) and the willingness/ motivation of native speakers to add content. In fact, when there is a will, there is a way: a good number of native projects have already started even in the absence of language-specific tools [Scannell 2013].

For digital vitalization (as opposed to digital heritage preservation, which we see as a fallback position) we must work together with speakers who are both motivated and literate. The body of text they produce constitutes the base (Stage 0) of the following language technology pyramid: 1. Locale or i18n support for the input and output of native characters; 2. Word-level tools (spellchecker, stemmer, dictionaries); 3. Phrase- and sentence-level tools; and 4. Speech and character recognition, machine translation. Besides Stage 1 capabilities, Stage 2 requires in-depth morphological analysis and generation (which will be trivial only for isolating languages). Stage 3 (POS taggers, named entity recognizers, chunkers) presuppose Stage 2 tools, and Stage 4, the peak of the language technology pyramid, presupposes all lower levels. Measuring the maturity of tools at the various stages, and creating them as needed, is the central task of digital language vitalization.

3. Conclusions

The Information for All Programme, and UNESCO in general, can foster the vitalization process by addressing the main issues directly. On the legal front, corpora, the lifeblood of modern computational linguistics, must be unencumbered by copyright, and IFAP/UNESCO can make sure that a research exemption is enshrined in the legal framework. At the national level, projects need to make their corpora not just searchable but also downloadable by ROAMing (randomize, omit, anonymize, mix). Both for international grants and those coming from national science foundations, linguistics should follow the lead of biosciences and demand, as a precondition of funding, open access to the materials collected. Finally, a wikipedia is a necessary but insufficient condition for digital ascent ("no wikipedia, no survival"), and digital communities (not just read-only material) are also needed. Therefore we suggest to give micro-grants to small communities (literary, theatrical, etc.) to document in their native language what they are doing.
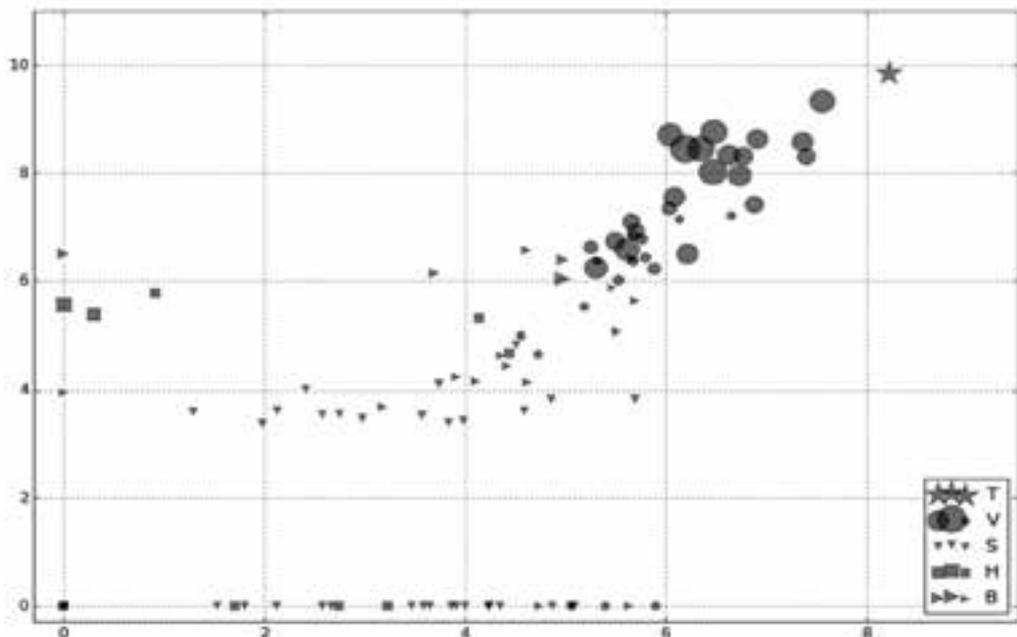
Figure 1. *Rough vitality assessment of languages in the Former Soviet Union*
(x: log population; y: log Wikipedia size; circle diameter: WP quality)

Table 1. Main vitality figures for languages in the Former Soviet Union

Language	SIL code	Vitality status	Population	Norm. WP size	WP quality
Abaza	abq	s	38,732	4,169	n/a
Abkhaz	abk	v	112,741	n/a	n/a
Adyghe	ady	b	491,801	431,288	n/a
Aghul	agx	b	22,677	43,811	n/a
Altay	alt	v	35,745	100,571	n/a
Alutor	alr	s	257	10,442	n/a
Armenian	hye	v	5,902,971	207,272,470	0.26
Avar	ava	v	761,961	1,696,067	0.11
Azerbaijani	aze	v	23,000,000	380,596,055	0.29
Bashkir	bak	v	1,221,341	36,074,594	0.29

Belarusian	bel	v	2,220,001	282,784,660	0.47
Buriat	bua	b	n/a	n/a	n/a
Chechen	che	v	1,361,001	13,900,569	0.01
Chukot	ckt	b	8,184	17,668	n/a
Chulym	clw	s	131	n/a	n/a
Chuvash	chv	v	1,077,421	22,366,496	0.15
Crimean Tatar	crh	v	475,541	2,295,102	0.06
Dargwa	dar	s	492,491	6,797	n/a
Dolgan	dlg	s	3,691	3,404	n/a
Dungan	dng	b	41,624	14,092	n/a
Enets	enf	s	33	n/a	n/a
Erzya	myv	v	336,315	1,041,262	0.08
Estonian	est	v	1,100,000	510,622,720	0.41
Even	eve	s	7,295	n/a	n/a
Gagauz	gag	v	178,024	4,238,180	0.14
Georgian	kat	v	4,237,711	215,886,780	0.33
Gilyak	niv	s	559	3,657	n/a
Gothic	got	h	n/a	369,638	0.14
Ingrian	izh	s	374	n/a	n/a
Ingush	inh	b	322,901	118,823	n/a
Itelmen	itl	s	133	4,191	n/a
Juhuri	jdt	s	17,156	n/a	n/a
Kabardian	kbd	v	1,628,501	3,187,050	0.31
Kalmyk	xal	b	291,794	750,654	0.03
Karachay-Balkar	krc	v	310,731	5,502,178	0.25
Karaim	kdr	s	94	2,431	n/a
Karakalpak	kaa	v	410,411	3,841,866	0.39
Karelian	krl	v	53,141	46,409	n/a

Kazakh	kaz	v	8,077,771	431,343,123	0.29
Ket	ket	s	376	3,546	n/a
Khakas	kjh	s	31,903	68,969	n/a
Khanty	kca	s	9,581	2,739	n/a
Khinalugh	kjj	h	1,668	n/a	n/a
Kildin Sami	sjd	h	551	n/a	n/a
Komi	kom	b	n/a	3,216,590	0.09
Komi-Permyak	koi	b	93,543	2,472,803	0.09
Koryak	kpy	s	2,916	n/a	n/a
Krymchak	jct	h	13,627	206,329	n/a
Kumyk	kum	b	426,551	n/a	n/a
Kyrgyz	kir	v	2,941,931	105,618,112	0.51
Lak	lbe	v	153,171	334,573	0.05
Latgalian	ltg	v	200,001	1,740,184	0.35
Lezgi	lez	v	788,721	n/a	n/a
Lithuanian	lit	v	3,001,861	581,134,721	0.45
Livonian	liv	h	7	607,201	n/a
Mansi	mns	s	941	3,026	n/a
Mari	mhr	v	475,874	7,053,484	0.09
Mari	mrj	b	40,531	3,759,145	0.06
Mari (Russia)	chm	s	n/a	n/a	n/a
Mingrelian	xmf	v	500,001	8,274,826	0.21
Moksha	mdf	b	92,765	1,097,308	0.16
Nanai	gld	s	3,843	n/a	n/a
Nenets	yrk	h	27,393	48,920	n/a
Nganasan	nio	s	461	n/a	n/a
Nogai	nog	s	73,305	n/a	n/a
Northern Altai	atv	b	12,728	14,652	n/a

Old Church Slavonic	chu	h	1	242,749	0.12
Old Georgian	oge	b	n/a	9,011	n/a
Ossetian	oss	v	577,451	5,982,030	0.09
Russian	rus	t	167,332,231	7,019,024,883	0.56
Rusyn	rue	v	623,501	2,736,759	0.08
Rutul	rut	b	25,923	28,060	n/a
Samogitian	sgs	h	n/a	n/a	n/a
Selkup	sel	b	1,501	4,918	n/a
Shor	cjs	s	6,811	2,510	n/a
Shughni	sgh	s	71,588	6,740	n/a
Southern Yukaghir	yux	s	62	n/a	n/a
Standard Latvian	lvs	v	1,552,261	282,525,870	0.58
Svan	sva	s	17,171	n/a	n/a
Tabassaran	tab	s	113,529	n/a	n/a
Tajik	tgk	v	4,479,651	16,106,757	0.06
Talysh	tly	v	206,196	2,359,896	n/a
Tat	ttt	s	17,320	n/a	n/a
Tatar	tat	v	5,406,111	90,404,247	0.36
Ter Sami	sjt	s	18	3,987	n/a
Tindi	tin	s	4,440	n/a	n/a
Tsakhur	tkr	s	22,188	n/a	n/a
Tsez	ddo	s	9,986	n/a	n/a
Turkmen	tuk	v	7,560,561	26,082,541	0.23
Tuvinian	tyv	v	248,429	n/a	n/a
Udi	udi	s	5,464	13,144	n/a
Udmurt	udm	b	467,156	2,556,163	0.10
Ukrainian	ukr	v	36,048,891	2,168,400,162	0.40

Ukrainian Sign Language	ukl	s	n/a	n/a	n/a
Urum	uum	s	122,654	n/a	n/a
Uzbek	uzb	v	25,000,000	203,427,158	0.21
Veps	vep	b	4,917	1,398,616	0.08
Võro	vro	b	54,773	n/a	n/a
Votic	vot	h	49	n/a	n/a
Yaghnobi	yai	s	8,124	n/a	n/a
Yakut	sah	v	450,001	1,2642,821	0.20

References

1. Comrie, B. (1981). *The Languages of the Soviet Union*. Cambridge University Press.

2. Kornai, A. (2013). Digital language death. *PloS ONE*, 8(10): DOI 10.1371/journal.pone.0077056.

3. Scannell, K. P. (2013). Indigeneous tweets, indigeneous blogs (website).

Daniel PIMIENTA

Director, Networks and Development Foundation FUNREDES
(Santo Domingo, Dominican Republic)

Daniel PRADO

Executive Secretary,
MAAYA World Network for Linguistic Diversity
(Buenos Aires, Argentina)

Exploring the Status of Languages of France on the Internet: Methods and Reflection of Possible Approaches for Other Groups of Languages

Abstract

The paper reports on recent studies which explore and analyze both the situation of French and a large subset of the languages spoken in France on the Internet and derives methodological lessons which could be useful in other countries for similar approaches directed to other group of languages. The content is the result of two different studies which were conducted by MAAYA[26] in 2013, the first one concerning the situation of French language on the Internet, funded by OIF[27]; the second one concerning the situation of the languages of France on the Internet, funded by DGLFF[28] of Ministry of Culture of France.

Introduction

MAAYA, either directly or through some of its members (FUNREDES[29], LOP[30] or Union Latine[31]), has conducted a number of studies to analyze the role of languages on the Internet, since 1988. In particular, a specific measurement methodology for a group of languages[32] in different spaces of the Internet

---

[26] *World Network for Linguistic Diversity: http://maaya.org.*

[27] *International Organization of la Francophonie: http://www.francophonie.org.*

[28] *General Delegation to French Language and Languages of France: http://www.dglf.culture.gouv.fr/.*

[29] *Networks & Development Foundation: http://funredes.org.*

[30] *Language Observatory Project: http://gii2.nagaokaut.ac.jp/gii/blog/lopdiary.php.*

[31] *http://unilat.org/.*

[32] *Latin languages (Catalan, French, Italian, Portuguese, Romanian and Spanish) as well as English and German.*

has been developed by FUNREDES and Union Latine. This approach has allowed to realize a sustained set of measurement campaigns, between 1988 and 2008. LOP studied all the languages in most of the Internet top-level domains of Asia and Africa, with the intention of measuring the space of minority languages. LOP based its studies on the systematic crawling of web pages of the chosen domain and the application of an algorithm for recognizing languages while FUNREDES/Union Latine use the counting facilities of Search Engine with a sampling of words designed for comparisons between languages.

Since 2008, the method of FUNREDES/Union Latine was put on hold as a consequence of the evolution of search engines and there is no systematic measurement any longer from the LOP, leaving the field without the means to monitor developments, except for using less reliable sources.

To overcome this situation, a very ambitious research project (DILINET, http://dilinet.org) was designed by MAAYA, with the support of Union Latine, UNESCO and OIF, and defined by a consortium of strong research institutions. DILINET's goal was to receive funding from a call for proposals of the Research Programme Framework 7 of the European Union. The two successive attempts, in 2012 and 2013, did not give positive results. MAAYA is redefining the project with Qataris partners and looks for funding in 2014 from the Research Fund of Qatar. Pending to the success of the DILINET project, there is a long period of lack of precise information about the evolution of the place of languages on the Internet.

The studies whose methods are exposed in this paper represent a methodologically much less ambitious alternative, but in any case are likely to report to an acceptable level of language development in the most visible areas of the Internet. It helps to fill the space and the time before the arrival of the DILINET project, focusing on contents related to some applications as well as some targeted uses for specific languages (French and a group of 15 among the families of languages spoken across the French territory[33]).

The proposed approach intents to escape from a simple one-time result so as to allow some level of monitoring developments in the coming years. Different approaches have been developed for French, one of the important languages of the world, and of the Internet, on one hand, and, on the other hand, for a subset of the languages used on the French territory that can be considered "minority" and are commonly called "languages of France" in France.

---

[33] *Alsatian, Basque, Breton, Catalan, Corsican, Creoles, Frankish, Franco-Proven al, Futunan, languages of Mayotte, language of Oil, Kanak languages, Occitan, Tahitian and Wallisian.*

The proposed approaches for French, and the set of languages of France which have been studied[34] could inspire studies on other languages with a large number of speakers, such as French, or on languages which are used within a given territory but cannot boast a large number of speakers (and thus have a relatively low profile on the Internet).

> *This study brings together the complementary and synergistic experiences of two independent studies realized by MAAYA in 2013, the first sponsored by the OIF, concerning the place of French on the Internet and the second sponsored by DGLFF of Ministry of Culture of France, for a subset of the languages of France. The document has the permission of both institutions for this public disclosure.*

This paper presents the methodologies used in these studies with the intention that it can be taken or adapted in other linguistic areas. Another paper [Prado and Pimienta 2014] will present the results obtained by combining the methodologies deployed.

Background and Approach

Although a number of indicators may be identified about the presence of the French language in diplomacy, education, science, international organizations, language translation, language editing, and many other aspects [OIF 2010], when speaking of French presence in cyberspace one is still in doubt and lately a hypothetical **8th place** on the Internet is mentioned, with little awareness of the subject under discussion.

The will, seemingly simple, to measure the "presence" of the French language on the Internet shelters actually a permanent misunderstanding, which is the consequence of the scarcity of information on the subject and the cause of the discrepancies in the figures given by different sources. Two different indicators are commonly confused:

- The estimated percentage of **Internet users**, speakers of a given language;
- The estimated percentage of **Internet content** in a given language.

---

[34] *The number of languages spoken in France is quite large, their origins are diverse and the number of speakers can vary from several million to several tens; the study focused only on a subset of the languages that are likely to be present in cyberspace. So non-territorial languages (that is to say, those originating in territories other than those occupied by the French Republic today, as is the case of several immigrant languages such as Arabic) and territorial languages with fewer than 50,000 speakers were excluded, unless they were teaching languages.*

Measuring the number of francophone Internet users or web pages in French are fundamentally different matters, reflecting different realities that deserve different attention: the first measure is related to the **digital access divide** (i.e. the physical access to the Internet) and the second one – to the **digital content divide**, a divide much less understood but more decisive.

Measuring the number of speakers of a given language implies a completely different protocol then measuring the number of contents in that language. When found in newspapers or in some reports, the figures for the "presence of French", require an exercise to differentiate if one refers to the language of Internet users or to the language in which contents are provided on the Web. Thus, the claim that French is in the 8th place on the Internet (information widely touted in the media) only makes sense if it is specified that the 8th population of Internet users is francophone. It comes back in no way to say that French is the eighth in terms of content.

Percentage of Internet Users by Language

This data comes from the most consulted source: Internet World Stats[35]. This source, which is far from meeting the expectations of rigorous statistics, at least has the merit to exist and to be the only one to be updated[36] on the language spread of Internet users. Its methodology is to determine the main languages used in each country and to cross this information with data from the ITU (International Telecommunication Union) about the total number of Internet users in each country. However, the ITU data are produced by governments, which is not necessarily a criterion of reliability. Indeed, on the one hand some countries tend to inflate the figures given to the ITU to demonstrate the success of their efforts to fight against the digital divide. On the other hand, there is no mention of the methodology used by Internet World Stats for weighting language users. It would appear that the only criterion is the official language of each country. Internet World Stats also uses various marketing sources, with probably no common methodology. Finally, this study is limited to the top ten languages of Internet users, in contrast to the GlobalStats company that provided comparable figures before 2007 but then disappeared and now some of the historical data can only be retrieved using the "Wayback Machine" of the archive.org[37] site.

---

[35] *http://www.internetworldstats.com/stats7.htm.*

[36] *It should be noted anyway that the figures have not been updated since May 31, 2011, leaving a very serious void in the world of indicators for the presence of languages on the Internet.*

[37] *http://web.archive.org/web/20041019013615/www.global-reach.biz/globstats/index.php3.*

Then there is a large category of publications (usually marketing companies) where figures are published and no method is revealed. It is impossible to validate the results. This was the case of an Inktomi study that was launched in 2001 with a great marketing noise and included gross errors. For example, presenting the worldwide percentage of web pages in a limited number of languages, the total of these percentages were 100%.

More recently, the site "W3Techs – World Wide Web Technology Surveys", which presents itself as the source of the most reliable and most complete information on the uses of the Internet, has the advantage of differentiating language of Internet users and language content, which it deals specifically on the page "*Usage of content languages for websites*"[38]. It computes its statistics based on the data by Alexa[39], a company able to provide statistics about usage of the Web through a toolbar that a sample of Internet users agree to install on their browsers. With its toolbar, Alexa accounts for access to the most visited sites, and then performs a ranking of the 25 million most popular websites of the Web, knowing that the Web has almost 650 million sites, including 200 million considered active and without duplicates. W3Techs thus takes the first 10 million sites ranked by Alexa and determines which languages they are written in through an algorithm of language recognition. The remarkable news is that W3Techs updates daily its action which allows for time-series from the start date of the service, in June 2013[40].

Another item of interest in the work of W3Techs is the ability to cross over some data:

- http://w3techs.com/technologies/cross/top_level_domain/ content_language allows to cross domain names and language content (so 27% of sites in French would belong to the France top level domain .fr[41]);

- http://w3techs.com/technologies/cross/content_language/top_ level_domain establishes the reciprocal cross (so 92% of sites from the .fr domain would be in French);

---

[38] *http://w3techs.com/technologies/overview/content_language/all.*

[39] *http://alexa.com.*

[40] *See the series for French: http://w3techs.com/technologies/details/cl-fr-/all/all or the one about the languages with more than 0,1% presence: http://w3techs.com/technologies/history_overview/content_ language*

[41] *In 2007, the FUNREDES/Union Latine study was computing a value of a little over 26% for this indicator and a value of just over 57% for the percentage of French sites located in France (including those of .fr as well as generic domains as .com or .org).*

- http://w3techs.com/technologies/cross/content_language/ranking establishes the intersection of the rankings in Alexa and the language parameter (so 61% of sites ranked in the top 1000 would be in English).

The today W3Tech figure for English (54% of all web pages) is higher than that of the FUNREDES/Union Latine study of Romance languages in 2007 (44%) and much higher than what would be our today projection (around 34%). There are two likely reasons for obtaining figures for English well above the reality.

1. **The management of multilingualism**: Funredes/Union Latine or the LOP focused on web pages and allowed to measure, within each website, the language of each different page; while W3Tech language focuses on websites and records probably as English websites whose homepage features English even if the rest is written in other languages.

2. **The use of Alexa**: being installed voluntarily by the user, Alexa is a good instrument to measure what users browse. But it is still necessary that the tool be known and used in a balanced manner between the different regions of the planet in order to compare usages consistently. And for now, this is not the case. In addition, Alexa used to measure usage and not existence; pages which are not visited by Alexa users are not identified. Moreover, W3Techs only considers the top 10 million of most visited websites according to Alexa, i.e. $10/650^{42}$ = 1.5% of existing sites. The visited sites will therefore necessarily include the mainstream media and the most reputable commercial sites of different countries, especially Western countries, the United States leading, but probably not many science sites, local or smaller shops distribution of most countries in the world.

3. **Language recognition algorithms have a tendency to overestimate English.** For an overview of possible linguistic biases of this study, it should be noted that Czech would have more pages than Korean or that Chinese, with an online population almost 10 times larger than German or Russian, would have fewer pages.

In spite of its limitations, W3Techs represents the most attractive source of indicators available today and one have to accept with satisfaction the progress it represents.

---

[42] *Following http://news.netcraft.com/archives/2012/11/01/november-2012-web-server-survey.html 650 million websites would be active.*

Evolution of Search Engines since 2008

Since 2008, the variety of search engines has been on the decrease and the generic search engines remaining on the market (Google, Yahoo, Bing/Live Search, Ask, AOL, Lycos, Excite, Exalead, Teoma[43]) have all evolved in the same way:

- significant reduction of the percentage of the indexed part of the Web (over 80% to less than 10% of the total space);

- total loss of credibility of the published figures of the number of occurrences of a given keyword;

- increased "intelligence" of the search keyword that led to the loss of the association keyword/results (either by introduction of automatic translations, or by introducing synonyms or supposed orthography correction).

From 2008, and in amplified manner as time passed, the size of the Web has become uncontrollable and can be considered in practical terms[44], approaching infinity. This results in the inability, cost wise, to conduct a comprehensive systematic crawl of the entire Web[45] from the search engines at present, and leaving merely an estimated less than 5% of the total pages[46].

Together with the rise of Web 2.0 the nature of the Web has changed and static pages (simple HTML) has left more room for dynamic pages. In the same period, the Internet topology for languages has changed radically with the stabilization of the relative growth of the initially well-represented languages (Western languages, in particular) and the rise of Asian languages and more recently of Arabic. In parallel, the nature of content has evolved by reducing the proportion of text data and increasing the audio and especially video (by the end of 2012 video traffic accounted for over 50% of the total, with forecasted growth of this percentage[47]).

---

[43] *Google would have (following http://en.wikipedia.org/wiki/Web_search_engine) a little more than 80% of the market, with however a trend to lower since 2010.*

[44] *Which means the computing cost for systematic crawling.*

[45] *In 2008, the figure of 127 billion pages was provided by various sources (especially the search engine CUIL, now gone, which claimed to crawl the entire Web). See the webpage maintained by archive.org: http://web.archive.org/web/20100916001435/http://www.cuil.com/.*

[46] *If there is one area where lack of transparency is the rule, it is the size of the indexes. Apparently several tricks are used (especially not to explore all pages within a site) to hide this limitation which does not of course apply the same way to all languages and is a handicap for minority languages.*

[47] *"Cisco Visual Networking Index - Forecast and Methodology, 2010–2015", http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360_ns827_Networking_Solutions_White_Paper.html.*

Under these conditions, the percentage of pages of fixed text in a given language could remain an indicator of some significance but, faced with a more complex reality, one must create other indicators that better reflect this complexity, and accept to deal with partial elements of a mosaic, rather than dealing with limited integral indicators.

In 2010, Union Latine, in collaboration with FUNREDES, conduced a first attempt to grab the perception of complex reality from the dispersed state of languages in the Web elements collection. This first experience was not concluded by a specific publication given the difficulty to rely on a broad base of data and the approximation of most of the data collected; however, it has indirectly contributed to a number of publications[48].

The year 2010 work was taken in all language directions so as to detect signals of changes and explore the best-known applications (social networks, blogosphere, peer-to-peer, search engines, and of course VOIP[49], Wikipedia and YouTube). Often, in the absence of alternatives, the data was built **from the geographical origin of the traffic to those applications** (as opposed to working directly on languages). These principles which have helped to outline a different methodological approach were amplified and extended in this new study.

Methodology for French

The proposed methodology is based on and expanded from the 2010 approach and applied to a given language, French[50], which is compared to other major languages used on the Internet. In this new study the approach is extended to the largest possible number of recognized application areas. The approach from the geographic traffic which gathered data on linguistic communities that use a given application is supplemented by intense and systematic research about the linguistic content of those applications.

This approach is therefore based on a very important and open work of collecting information about the languages in Internet applications, compilation and organization of data, assessment and validation and direct crossing with different specific studies seriously and ultimately putting them

---

[48] *See [Pimienta 2012] and [ITU 2010].*

[49] *Voice over IP like, for example, Skype.*

[50] *This does not completely rule out to work with other languages because it will sometimes be necessary to offer comparisons.*

into perspective in order to trends and composite indicators that report on emerging developments[51].

The key methodological elements of this approach are:

- Considered spaces and applications;

- The selection and analysis of the sources of information on the place of French related to the selected spaces and applications;

- Demolinguistic data that are used to put into perspective the data collected;

- The attempt to bring together the scattered data into a synthesis that meaningfully informs on the place of French on the Internet and put in perspective.

Considered Spaces and Applications

Around one hundred applications and spaces of the Internet have been identified as likely to shed light on the role of language on the Internet. Thirty of them were unable to offer reliable and usable data evenly: these have been temporarily excluded and will be integrated as soon as data offering permits.

The chosen initial list (including those that were discarded) is presented and organized in the following tables, by application type or space.

Infrastructures	Online Books	Telephones & Tablets	Messaging & IP Telephone
Internet users per language	Virtual library	Smartphones	Skype
Computers per country	GoogleBooks	Tablets G3	QQ
Websites per population	Amazon	Data Sims	AIM
Websites per users	GoogleScholar	3G	ICQ
Internet penetration			Yahoo!
High bandwidth penetration			
Automatic translation traffic			
Online linguistic tools			

---

[51] *To give an order of magnitude, the 2010 work allowed to reference some thirty links (eg. http:// socialmediastatistics.wikidot.com/ which allowed to know the distribution of "tweets" or "Facebook" pages per language). The goal here was to collect, compare and evaluate at least a hundred links for a potential source so to expand coverage. In fact, hundreds of possible sources were evaluated and nearly 200 of them were selected, analyzed and classified.*

File Download & P2P	Social Networks	Blogs	Webpages Counting
Megaupload	Wikipedia	Technorati	W3 Counters
Rapidshare	Facebook	Blogs	WebBoar
Filefactory	Twitter	WordPress	Internet Archive
Depositfiles	LinkedIn	Google Blogs	Google
Hotfile	Viadeo	Blogger	Baidu
Uploading	Xing	Blogspot	Wolfram Alpha
Uploaded	Yahoo	Sina Weibo	MSN
Fileserve	Google+	Technorati	Bing
Mediafire	Windows Live Profile		Foofind
Gigasize	MySpace		Rtbot
Bitshare	Livejournal		CC Search
4shared	Secondlife		Altavista
	Ning		Yandex
	Tuenti		Wikia Search
	Hi5 Orkut		Open Directory Project (DMOZ)
	Badoo		
	Tumblr		
	Instagram		
	Sonico		
	Qzone		
	YouTube		
	Googleplay		

Email	Search Engines	Browsers	Operating Systems and Application Suite
Gmail	Google	Chrome	Windows
Hotmail	Bing	Firefox	Linux
Yahoo	Yahoo!	IE	Mac
Yandex Mail	Yandex	Safari	Ios
Icloud	Baidu	Opera	Android
Outlook	Others	Others	Openoffice
			Microsoft Office
			Others

Analyzed Sources

A significant effort to collect sources on the presence of languages on the Internet, in general, and on French in particular was deployed. This chapter is dedicated to the sources that fed the results presented.

Considerations about Sources

The lack of production from FUNREDES and LOP has given way to a place where the vast majority of sources are companies of business intelligence or online marketing which filter some partial information for free (and usually without revealing the methodology) as a means of promoting their paid services. Apart from the traditional and reliable statistical sources that provide useful information to work on languages (UN, UNESCO, ITU, OECD, EU, etc.), while having the same difficulties in identifying the issue of the languages on the Internet, some consultants or experts who are dedicated to gather as much information about a space or a given application, share their results in a website in order to promote their expertise. Analysis of those professional sites is often very useful. Gathering this set of broad but imperfect information allows, with method and with some difficulty, to take stock segmenting the problem following applications (Wikipedia, Twitter, YouTube, Facebook, ...) or space (search engines, email, etc.).

Another limitation to consider however in this type of analysis is that the degree of globalization of spaces and applications is becoming more variable and, increasingly, countries or regions adopt their specific applications to the detriment of the major world applications (such as Baidu instead of Google in

China and Vkontakte and Odnoklassniki instead of Facebook in Russia). We must take this factor into account when quantitative data on the use of languages (or countries) are established for an application or a given space. Thus, the conclusion that in the non-professional social networks, French, for example, is the fourth or the first language from its respective results in Facebook or Viadeo, does not make sense to the extent that any of these applications knows even penetration across all geographic areas and thus language. To obtain a credible indicator about the ranking of French each application of non-professional social network type should be weighted based on its distribution in the world (total weight and relative presence in a country), a task that is beyond the reach of this study.

A simple overview would emphasize that the most globalized applications are Wikipedia, Twitter and YouTube[52], those which are relatively globalized are Facebook and Google with exceptions for some countries (China, Russia, Kazakhstan, Korea, etc. which used local search engines) or show different habits (Yahoo! is much more used than Google in Japan).

For most other applications and spaces one needs to be careful and trade-off the linguistic conclusions from the results to possible biases in usages.

Source Selection

Since 2010, several hundreds of potential sources have been analyzed and stored and constant monitoring is performed by search keywords or external links consultations of the sources. In total, some hundreds of sources have been detected and analyzed.

Many of them were excluded for one of the following reasons:

- The field of study was too small or partial.

- They appeared too biased.

- The numbers were not updated or there were significant differences in dates inside mentioned sources.

- The methodology used did not allow to compare the data.

- The methodology used did not seem relevant, adequate or credible.

- There were doubts about the reliability of the source.

---

[52] *What is less the case for YouTube since it has a competitor in second position, Dailymotion, for which France Telecom is the majority shareholder, and which has a strong francophone presence, even if its name is not indicated, and that it may be tomorrow taken over by Yahoo.*

Around 200 sources of information (URL, articles, books or other) that could identify indicators of the presence of languages in different areas were finally selected, classified, evaluated and rated.

Source Classification

Each element of this sample was rated between 10 (excellent) to 5 (average); those scoring less were automatically rejected. The rates have been prepared based on several criteria: relevance, confidence, reach, transparency of method, etc.

The rating results are the following:

Rate	10	9	8	7	6	5	<5
Number	6	4	19	37	38	25	59

59 sources rated less than 5 have been kept for later evaluation.

For each source the following parameters were recorded:

- The last year of publication.
- Target (global, Francophonie, etc.).
- If the source is updated frequently or not.
- Type of source (eg. meta-information).
- Area of application of the source (eg. Facebook).
- If it is specific to the language or not.

The sources will be exhibited in the future in a clearinghouse, a kind of database of web links, with the corresponding parameters in order to maintain an observatory. Meanwhile, the degree of rapid obsolescence of sources and the dynamic creation / removal of Internet pages is such that it is not appropriate in this article to mention all sources.

Demographics and Demolinguistic Data

Really monolingual countries are an exception; multilingualism is rather a rule, such as, for example, in France, the United States, China or Cameroon (one of the countries with the greatest linguistic diversity in the world), or even in micro-states like Monaco, Malta or Singapore.

In almost every country in the world counting speakers of different languages must be undertaken if one wishes to transform data by country (which are natural sources in most cases) into data per language (which are those needed for this study).

Most studies, and it is a weakness for their linguistic theme, provide data by country or region and tend to extrapolate the results to the official language of each country[53], where details of the language are set. Such a simplification carries important errors[54].

To compare French to other languages, it is necessary to establish reliable statistics of all languages of the world, or at least those that are spoken by a large majority. Reliable statistics exist for some languages spoken in well-defined territories and knowing an important development, provided also that they have an official status or some sort of protection and their diaspora still speaking the language is well studied. This is the case in many languages spoken in Europe, for example, Italian, Lithuanian, Polish and Catalan. But this task becomes complicated for languages that meet none of those conditions.

Difficulties with Languages without Supervision

To take examples that are meaningful to the French public, in the case of languages like Occitan or Franco-Provençal, for which, although spoken in developed areas and on fairly well defined territories, lack of institution of guardianship affects the quality of the figures for the number of speakers.

Difficulties with Languages Occupying Large Territories

For some languages occupying extensive territories and diverse socio-economic conditions, there are relatively reliable statistics; This is the case of French or Spanish, for example, through OIF and the Cervantes Institute respectively.

But for other languages with similar characteristics, such as English, Chinese or Portuguese[55], there are very important differences between sources, particularly on "L2" speakers[56].

Demolinguistic Conflicting Sources

Divergent methodologies in counting speakers in the many demolinguistic studies add up to the above mentioned complications. To be able to compare

---

[53] *One of the rare counterexample is against Wikipedia that provides remarkable linguistic data http://meta.wikimedia.org/wiki/List_of_Wikipedias/sortable.*

[54] *Spreading the data proportionally of the number of speakers is another simplification of course (as the digital divide is not evenly spread in the population and migrants often have less access to the digital world; however the errors produced by this simplification are by an order of magnitude lower.*

[55] *Camões Institute, however, starts to propose figures with an acceptable margin of error.*

[56] *L1 is the notation used for mother tongue speakers. L2 is the notation used for speakers controlling or using fluently one language but not having it as their mother tongue.*

all available studies on the number of speakers of all languages in the world to create a comparative "super-grid" is a huge task, above the resource of this study. Thus, it is unthinkable to parallel the results of different studies without detailed analysis because it would put on the same level figures obtained by non-comparable definitions or inhomogeneous methods.

Language Typology

An additional problem is to be solved on typology of languages. To come back to the example of Occitan the question is: should speakers of Gascon be accounted for speakers of Occitan language or independently of? If so, should then Provençal, Auvergnat, Languedocian, etc. also be accounted for separately? Some approve more, others less. Should the German language be taken as a whole or as a set of various dialects, sometimes very distant from each other? What to do with the Arabic language, also characterised by a large dispersion of dialects? Should we consider literary – or classical – Arabic only or all of dialectal Arabic (Algerian Arabic, Levantine Arabic, Egyptian Arabic, etc.)? Should Calabrian be associated with Neapolitan as Ethnologue suggests or only the northern Calabrian, attaching the southern Calabrian to Sicilian as Wikipedia suggests?

Second Language Speakers (L2)

But without doubt, the most acute dilemma is how to take into account "L2" speakers for vehicular languages. These people do not have that language as their mother tongue, but master it enough for common use. If for some languages the number of speakers may not have a significant impact on this study[57], it is clear that L2 speakers of English, French, Spanish or Russian influence greatly the results. Even if there is no L1 speakers (native) of English in Ghana, according to Ethnologue, the L2 one million speakers of English in this country seems to produce more content on the Internet than the 2.5/3 million speakers of Ewe living in this country. Also, if the population of Paraguay is mostly guarani-speaking, fluency in Spanish (L2) by the general population relegated Guarani on the Internet behind the Spanish language.

But there are many cases where more than one language have a vehicular role for certain populations (English and Swahili in East Africa, English and Hindi in India, French and Hausa in francophone Africa, etc.). In this case, what is the preferred vehicular medium for using the Internet? Hausa and Swahili

---

[57] *As it relates to technologies that are not always in use for some populations of less developed countries, such as Quechua or Swahili, also used as L2 language.*

seem to be relegated behind French and English, following LOP studies and [Diki-Kidiri 2007], but it would be less and less the case for Hindi, for example.

Although there are some reliable data on L2 speakers of French or Spanish in countries where these languages are official (de jure or de facto), it is impossible to have statistics both accurate and comparable with each other on the use of major languages of communication[58] other than in countries where these languages are official. However, it is common to see websites in English or having one of the versions in English in many countries where English is not an official language. Also many websites are in French (or include pages in French) in Spain, Germany or the United States, so much so that previous studies from FUNREDES/Union Latine had shown that Spain or the United States produced more French contents than the whole countries in Francophone Africa.

Thus, given the complexity of the panorama, some choices have been made that are not entirely satisfactory, but allow at least to achieve consistent and measurable results.

Demolinguistic Choices

To take the best account of all these elements a number of choices have been made for the best trade-off between homogeneity and reliability of demolinguistics data.

1) Ethnologue for L1

When accounting for mother-tongue speakers, the choice was naturally made on Ethnologue[59], the only source providing continuously updated figures about all the languages of the world. This source is often inaccurate in its figures (this is precisely the case of the French language), it is often incomplete and its updates are not homogeneous, but it is the only one to provide dynamic data on all languages of the world by applying a relatively consistent methodology. The resolution exposed and explained in 2010[60] by the authors of Calvet Barometer[61] inspired that decision.

---

[58] *English, French, Spanish, Russian, Chinese, Portuguese, Arabic, etc.*

[59] *Ethnologue announced a new online version with many changes of its statistics after we finished the study. It must remain clear that demolinguistic data from Ethnologue taken in this study are those published in May 2013.*

[60] *http://wikilf.culture.fr/barometre2012/tmpl.php?data=doc/methodologie/index.*

[61] *http://wikilf.culture.fr/barometre2012/.*

Wikipedia was an option, but the famous encyclopedia does not offer a stable methodology for counting speakers (and for good reason, since the nature of Wikipedia is not to provide centralized information, each item being independent of the others). While Wikipedia is taking Ethnologue as its primary source for most languages inventoried, it does take as sources, for other languages, separate studies offering different methodologies and therefore not comparable with each other. Moreover, for some languages, Wikipedia is cautious and gives ranges and not precise figures[62].

The case of Albanian, to mention a language that can have more accurate data due to the age of the studies that have been dedicated to it, is enlightening. Ethnologue mentions 15,000 speakers in Turkey while the Wikipedia article talks about three million speakers in this country while including Ethnologue among the sources!

The communities of language professionals are often very critical of Ethnologue and this choice is obviously not without its problems[63]; however, the inconveniences affect only marginally this study because they mostly apply on languages identified in "other languages" statistics.

2) Different but reliable sources for L2

Regarding L2 speakers, Ethnologue is not at all satisfactory for some languages, including English, French, Spanish and Portuguese (for which there are reliable data otherwise); it will only serve as a reference where no better source is available. Thus, the following sources which are more reliable (and detailed) were used:

- For English: Wikipedia – List of Countries by English speaking population International[64].

- For French: Organization of La Francophonie[65].

- For German: National Geographic Collegiate Atlas of the World. Willard, Ohio: R.R Donnelley & Sons Company. April 2006, p. 257–299, cited in many sources including the article on the German Wikipedia (English version)[66].

---

[62] *This is notably the situation of English where 309 < L1 < 380 millions.*

[63] *Many problems have as a matter of fact been identified and noted during the process of this study.*

[64] *http://en.wikipedia.org/wiki/List_of_countries_by_English-speaking_population.*

[65] *http://www.francophonie.org/IMG/pdf/Synthese-Langue-Francaise-2010.pdf.*

[66] *http://en.wikipedia.org/wiki/German_language.*

- For Portuguese: Comunidade dos paises de língua portuguesa (R & D Unit)[67].
- For Spanish: "El español en el mundo", Report from Instituto Cervantes[68].

It should be noted that for the Arabic language, Ethnologue provides figures in a confused manner leading to an impressive number of articles that reflected an erroneous oversizing of the number of speakers. Indeed, in view of the number of speakers of Arabic dialects mentioned by Ethnologue (206 million[69]) and classical Arabic L2 (246 million) many articles consider the total number of Arabic (L1+L2) as close to 450 million by putting together logically the two numbers. As a matter of fact, there is an overlap as many dialectal Arabic speakers are also counted with classical Arabic as L2. As many papers report an additional number between 20 and 80 million Arabic mother tongue speakers, without mentioning specific or searchable source, the Ethnologue statistics were used, but interpreted as the number of speakers of Arabic L2 the subtraction of two values (40 million or so) and therefore the total number of Arabic L1 + L2 is identical to that of classical Arabic L2.

3) Wikipedia for demographics data

After studying various sources on the figures for the demographics of all countries of the world, French Wikipedia[70], which seems to federate the best and well updated sources, was used. Only a small problem (surmountable) remains to deal with: the current version, unlike past consulted versions, makes no demographics discrimination about non-metropolitan territories of France as it did for the United Kingdom, the United States, Norway, the Netherlands, China, etc. For comparative purposes the calculations were made as these overseas regions are separately included in most studies on the tools of the Internet.

4) Results window for some segments

Two different simulations were presented for specific areas or applications, one that discriminates by language, and one that takes into account a wide use of the main working language. Consequently, there are no exact figures about

---

[67] *http://www.idcplp.net/?idc=30&idi=5623.*

[68] *http://cvc.cervantes.es/lengua/anuario/anuario_12/i_cervantes/p01.htm.*

[69] *This figure is already confusing because by adding the figures by country, the total L1 speakers of all dialects of Arabic is 220 million.*

[70] *http://fr.wikipedia.org/wiki/Liste_des_pays_par_population.*

the use of a given application, but a range between two values indicating the relative ranks of the French language over other languages and the relative percentages of contents or use.

5) A case by case approach for specific situations

Arbitrary choices had to be made on specific cases because sometimes some studies took into account the macro-language while others dealt with only one of the languages of the family. It is the case of German (covering languages differentiated by the proposed standard German typology of Ethnologue: Bavarian Frankish Main, Swiss German, etc.), Arabic, Chinese, Pashto, Persian, etc. In this case the macro language has been taken as a sole reference in order to permit comparison. Thus, when we speak of Chinese, it will be all Chinese languages (Mandarin, Hakka, Yue, etc.), the same is for Arabic, German, Malay, Fulani and other groups.

Although these considerations have a very small impact on the results of this study which focuses on French, they are reported as from a critical standpoint those choices could appeared opposed to sound logic processing of language variants.

Global Assessment Method

After the presentation of so scattered records from French on the Internet (either as a mother tongue, L1 or a mother tongue and a second language, L1+L2), a question arises of the ability to extract statistically meaningful global vision from those fifty rankings in different spaces or applications. Is there a plausible way to give a meaningful and comprehensive outcome for the place of French on the Internet?

It seems clear that a simple average of rankings between different spaces and applications has little meaning for L1 as well as for L1+L2. One possibility is to weight various classifications according to the relative importance of the corresponding space or application and to obtain a weighted average that provides some meaning to an overall estimate. A final possibility is to establish a series of qualification parameters of each outcome based on elements of credibility of the result and the average weighted according to the value placed on these parameters.

All three methods have been deployed for the purpose of comparison, and to enable the development of a global ranking to integrate all the results and reflect with some accuracy the place of French on the Internet.

The rankings obtained are presented below, sorted by ascending values for L1+L2 and L1 elements evaluated. The classification of French as a mother tongue (L1) is between 4 and 12 and that of French as a first and second language (L1 + L2), varies from 1 to 8.

A **simple average of these rankings is respectively 6.8, for L1, and 4.2, for L1 + L2**.

A simple weighting is to assign a weight between 0 and 10 (marked P in the table presenting the results) for each item to show the relative importance of each space or application (thus a maximum weight of 10 is assigned to the elements "language of Internet users" and "percentage of pages in French" and a weight of 3 to Hi5 and CCSearch applications). Despite its subjectivity this method allows to come closer to the objective of overall weighting.

With the values proposed **the weighted average would be 7.4 for L1 and 4.3 for L1 + L2**.

A slightly more complicated weighting, which could translate more accurately the importance of the parameters, is available with this equation to calculate I, the result value as an indicator of French on the Internet: I = AxBxCxD / 1000 where:

A = Degree of globalization of the parameter (0 to 10)

B = Degree of reliability obtained for the values of this parameter (0 to 10)

C = Confidence level for the data obtained for French (0 to 10)

D = Relevance of the parameter for French (0 to 10)

This index is applied to the values L1 and L1 + L2 (noted L12 in the Table hereinafter sorted by growing values of L1 and L1 + L2):

Element	A	B	C	D	I	L1	L12	P	L1xI	L12 xI	L1 xP	L12 xP	TYPE
Viadeo	2	5	7	10	7		1	6	0	7	0	6	**SN**
Tumblr	6	6	7	6	15	4	2	4	60	30	16	8	**SN**
Hotmail	5	5	6	6	9		2	4	0	18	0	8	**APP**
Open Office	9	9	9	8	58		2	5	0	117	0	10	**APP**
Blogs.com	6	7	7	5	15		2	5	0	29	0	10	**BLOG**
Open directory	9	10	7	9	57		2	7	0	113	0	14	**CONTENT**
Badoo	6	5	7	5	11	5	3	3	53	32	15	9	**SN**
Foofind	7	7	7	6	21	5	3	3	103	62	15	9	**APP**

Smartphones	10	6	8	9	43	7	3	4	302	130	28	12	**INFRA**
Servers / hub	10	9	9	7	57	8	3	5	454	170	40	15	**INFRA**
3g	10	6	9	6	32	9	3	3	292	97	27	9	**INFRA**
Amazon	7	9	9	8	45		3	6	0	136	0	18	**LIVRES**
Gmail	7	5	8	6	17		3	6	0	50	0	18	**APP**
Yahoo	5	5	6	6	9		3	4	0	27	0	12	**APP**
Facebook	8	7	7	6	24	5	4	7	118	94	35	28	**SN**
Twitter	9	8	7	9	45	6	4	8	272	181	48	32	**SN**
Livejournal	8	7	7	7	27	6	4	5	165	110	30	20	**BLOG**
LinkedIn	7	7	7	7	24	7	4	6	168	96	42	24	**SN**
Internet World Stats	10	6	10	10	60	9	4	10	540	240	90	40	**USERS**
Gigasize	7	7	7	6	21		4	4	0	82	0	16	**P2P**
Windows Live Profile	7	7	7	6	21	5	5	4	103	103	20	20	**SN**
Instagram	7	5	7	6	15	6	5	5	88	74	30	25	**SN**
Google +	8	7	7	6	24	7	5	7	165	118	49	35	**SN**
Skype	8	7	7	8	31	7	5	7	220	157	49	35	**APP**
Hi5	7	6	7	4	12	8	5	3	94	59	24	15	**SN**
Internet Archive	9	7	7	9	40	8	5	7	318	198	56	35	**CONTENT**
CCSearch	6	7	7	6	18	8	5	3	141	88	24	15	**APP**
Wikipedia	9	10	10	10	90		5	8	0	450	0	40	**CONTENT**
YouTube	8	7	7	8	31	7	6	7	220	188	49	42	**VIDEO**
Icq	5	7	7	5	12	10	6	3	123	74	30	18	**APP**
W3tech	10	7	10	10	70		6	10	0	420	0	60	**PAGES**
Orkut	2	5	7	3	2		6	6	0	13	0	36	**SN**
Ixquick	6	7	7	6	18		6	3	0	106	0	18	**APP**
Bitshare ++	7	7	7	6	21		7	4	0	144	0	28	**P2P**
Mobile Telephony	10	9	9	7	57	12	8	3	680	454	36	24	**INFRA**
Rapidshare	7	7	7	6	21		8	4	0	165	0	32	**P2P**
HighBandwidth	10	9	9	7	57	5		4	284	0	20	0	**INFRA**
Aol/Aim	5	7	7	6	15	5		3	74	0	15	0	**APP**
Ning	7	7	7	8	27	6		5	165	0	30	0	**SN**
Msn	7	7	7	6	21	6		5	123	0	30	0	**APP**
Wordpress	8	7	7	7	27	7		5	192	0	35	0	**BLOG**
Average						6.8	4.2		7.4	4.3	7.2	4.2	

The values obtained for the average are as follows.

	Simple Average	Simple Weighted Average	Multi-criteria Weighted Average
L1	6.85	7.18	7.44
L1+ L2	4.22	4.21	4.30

It is therefore reasonable to conclude, based on the parameters established and the results gathered, that the general ranking of French on the Internet, all criteria included, is **between the 7th and 8th for L1**, and **between the 4th and 5th, rather close to the 4th, for L1+L2**.

The comparison of these results with the respective ranking of French in the world, in different areas, suggests that French on the Internet is well above its ranks in the world, based on the number of speakers although there are other areas where its ranking is even better (as its presence in international organizations, the production of literature and translation).

To complete the statistical treatment of results, it may be interesting to understand what are the spaces or applications where French ranks better:

Type	L1	L1+L2	
Books		3	*
Blogs	6,5	3,3	
Applications	6,7	3,6	
Social Networks	7	4	
Infrastructures	7,9	4	
Users	9	4	*
Contents	8	4,1	
Video	7	6	*
P2P		6,3	

(*) Only one result for this type

The two least favorable rankings for L1 are those of Internet users and infrastructure, which seems quite logical as they reflect the digital divide that affects a significant part of the Francophone world, specifically in Africa.

Regarding the rankings L1+L2, which are in principle the most significant, it should be noted that the less favorable ranking is related to peer-to-peer and video, however, those are areas of rapid growth of the Internet.

Regular monitoring of all these results would allow for an observation of the place of French on the Internet, which is very useful to determine the policies to be followed to support and enhance its presence.

Methodology for Languages of France

Introduction

Selection

The study, for budgetary reasons, is not planning to include the whole range of languages spoken on the French territory, in the broad sense of the term. It was then necessary to establish a selection with the objective to focus on languages which are most likely to be present on the Internet, excluding immigrant languages that are often not a minority (such as Arabic or Portuguese). **The selection criterion was the following: a local language holds more than 50,000 speakers or is an officially recognized language of education**.

This leads to the following list of languages and language families (*), in alphabetical order: **Alsatian, Basque, Breton, Catalan, Corsican, Creole (*), Frankish, Franco-Provençal, Futunan, Kanak languages (*), Mayotte languages (*), Oil language, Occitan (*), Tahitian, Wallisian**.

Method

If it is certain that a number of methodological elements which have been used for the part of the study focusing on French will be useful again for other languages of France, the method used for French, a language that combines several tens of millions of speakers in the world and represents more than 4% of the total web content, cannot be applied equally for languages that weigh on the Internet 10 times less (as Catalan) or 100, 1000, or more likely 10,000 times less (like Kanak languages). The reason is simple: there is virtually no Internet resources to provide data to allow a quantification of the presence of the selected languages on the Internet[71].

---

[71] *Except perhaps for Catalan whose Spanish side has long relied on a significant Internet presence and was the first to create a linguistic top-level domain on the Internet: .cat existed since 2005 and now has more than 50,000 sub-domains.*

Consequently, the approach must be both less ambitious in the scope of the search, and at the same time more systematic in the depth of researching resources. Looking for sites that provide information about the quantified presence of a language in a given application or a given space of the Internet is not enough. It is now time to collect, relatively close to the completeness, all resources on the Internet that provide information on the presence of this language in cyberspace. If within the sampling quantified presence of the language in some areas or applications of the Internet is obtained, it is good! But it is not reasonable to give this as a goal.

Although this compilation of sites related to a language cannot be fully comprehensive[72] it will, if well informed and well organized, allow to draw useful lessons about the state of the language, which comparisons between languages could permit to refine and put into perspective. Besides, organizing these references will lay the foundation for a systematic monitoring and observation of further developments.

The real issues are the following:

- What are the *edges and boundaries* of the definition of sites related to a particular language?

- What are the *parameters* that need to be informed enabling a statistical post-processing of this collection of sites so as to make little sense emerge?

- What is the *method* used for the census and the creation of this collection?

- What happens to the language *families* in this research?

Resources Selection Criteria

The sites chosen should be related (directly or indirectly) with the language (not with the territory). For example, it is not about Corsica but the Corsican language. Purely tourism oriented sites are not kept, except to make a significant contribution for the language. Purely cultural sites are kept if they are related to the language (poetry or drama in regional language for example). Only articles or books available online for free are referenced. In case a site reproduced the contents of another site, the original source will be systematically sought and preferred.

---

[72] *How could it be anyway in a virtual universe that never stops evolving at breakneck speed? In one week a non-zero percentage of referenced sites will be gone and another percentage of new sites will have appeared.*

So the best (but unfortunately few) references are websites, books, scientific papers, or even presentations about the presence of a selected language on the Internet and/or providing data about that presence.

Good references are found in the list below:

- Meta-references about language (database, umbrella organization around that language, clearinghouse, etc.);

- Linguistic resources (dictionaries, translators, etc.);

- Discussions about the language;

- Cultural references with a direct relationship with the language;

- Serious and free offers of language courses online;

- Blogs in or about the language.

Testing and Census

A step by step algorithm was adopted to establish the census in each language:

1. Simple search, with the most common name of the language and "Internet" to find the main sites, declining the first 100 pages.

2. Analysis and processing of identified sites/pages and note-taking of associated external links if they exist and have a certain level of relevance.

3. Listing websites or web pages from associated links and back to the previous point. The loop stops when it is clear that one always falls back on the same sites or pages of external links.

4. More sophisticated searches to complete the process, with the same method for external link: GoogleScholar, books, blogs, other terminologies, English terminology, other search engines.

This method, applied intensively but in a very limited time (a few days per language), allows to draw a realistic picture of what exists. However it should be clear, especially in the current state of search engines, that the final sample is not exhaustive although the ambition is to approach 80% of sites that match the search criteria. Many quality sites that are not adequately referenced may get excluded of this effort[73].

---

[73] *This is why an effort to develop a clearinghouse should be accompanied by the ability to allow players to bring their own reference or suggestions of other sites.*

This method will of course reap repetitively for different languages, sites or pages with general information on all the world's languages or general data on all (or a subset) of the languages of France. These items will be grouped into two additional categories: "general" or "languages of France" and references will not be repeated in the list corresponding to each language sites.

The recommended method to whoever wants to use the clearinghouse to gather information about one of the languages of the study is to start by looking for that language within the sites listed "General" and then to those listed "Languages of France", before continuing with the examination of the resources recorded under the heading corresponding to that language.

Parameters Kept

- URL;
- Description;
- Year;
- Update (yes / no);
- Sector: Government, academia, non profit organizations, personal, business;
- Type: scientific paper, database, library, blog, meta-information, portal, social network, language resource;
- Language: local language, French, German, English, Spanish or several;
- Quantitative data (yes / no);
- Comments.

Evaluation

It is not about a value judgment about the resource! The ratings are allocated following the level of contribution (or proximity) to the theme "presence of the language on the Internet".

Thus, the following ratings have been agreed on:

9: Outstanding contribution to the designated topic or providing meaningful data.

8: Strong contribution or interesting data.

7: Interesting contribution or original data.

6: Average contribution.

5: Indirect relationship with the theme.

4: Indirect relationship with the theme but with little content.

3: Not accessible but retained because of special interest.

<3: Dismissed.

Statistics

The process brought together more than 1,000 references with a number of references by language ranging from just over 30 (Wallisian) to just under 250 (Occitan), with 10% of those with a score greater than 8.

From this material it was possible to derive a number of interesting statistical results that provide information on the situation of the languages studied on the Internet.

The rate of invalid external links (return code 404) informs about the vitality of the language on the Internet (a value greater than 20%, as in the case of Creole, is a symptom of the language having problems, as opposed to a minimal value in the case of Corsican, which shows a very strong vitality).

The breakdown of results by type offers interesting information, as seen in this table which presents a subset of languages:

Types	Gen.	LDF	Breton	Cors.	Creoles	Franco-Provençal	Kanak	Occitan	Tahitian	TOTAL
Publications	23%	40%	15%	22%	25%	21%	40%	23%	13%	**24%**
Data Base	4%	2%	0%	4%	0%	7%	2%	2%	4%	**3%**
Blogs	0%	2%	6%	22%	2%	6%	8%	15%	0%	**9%**
Media	0%	2%	0%	2%	1%	2%	0%	3%	0%	**1%**
Meta	14%	7%	16%	2%	10%	2%	5%	2%	19%	**9%**
Portal	10%	10%	44%	24%	28%	25%	14%	24%	30%	**23%**
Linguistic Resource	48%	38%	18%	23%	31%	28%	26%	30%	28%	**29%**
Social Network	1%	0%	1%	0%	2%	10%	3%	0%	4%	**2%**
Total	**7%**	**6%**	**8%**	**9%**	**10%**	**10%**	**12%**	**24%**	**4%**	**100%**

The breakdown of results by sector can characterize somehow the language situation on the Internet. The following table, which focuses on a subset of languages is quite telling:

	Org	Edu	Per	Gov	Com	Other
General	*27%*	**49%**	0%	8%	7%	8%
Languages of France	20%	48%	7%	**23%**	*2%*	0%
Breton	**52%**	17%	6%	*3%*	22%	0%
Corsican	15%	24%	**27%**	19%	14%	0%
Creoles	24%	31%	14%	5%	**26%**	0%
Kanak	21%	48%	12%	*7%*	*6%*	6%
Occitan	39%	19%	25%	*7%*	9%	1%
TOTAL	**31%**	**30%**	**17%**	**8%**	**12%**	**2%**

Resources classified as "**General**" come mainly from the academic world, followed by the voluntary sector (in this case often global associations). Resources classified as "**languages of France**" also come from academia, but this time the statistics shows the effort made by the government sector which is second close to the voluntary sector. More than 50% of resources in **Breton** from a vibrant voluntary sector but tourism play a role in propelling the commercial sector in the second place. **Corsican** shows a very balanced result in terms of sectors, evidence of a dynamism shared by all sectors, the highest score in terms of personal pages and the second place in terms of government (local government in this case) which is quite significant. The presence of **Creole** on the Internet is dominated by the university, tourism again shows its presence and the local government could do much better! The difference between **Breton** and **Occitan** is in the power of personal pages for the second.

It is also possible to exploit the statistical parameter on the use of languages:

	Average	Group with higher %	Group with lower %
% of sites in English	10%	General	Occitan
% of sites in French	48%	Languages of France	Tahitian
% of sites in local languages	7%	Corsican	Languages of France
% of sites in French & local languages	19%	Breton & Corsican	Languages of France
% of multilingual sites	18%	Tahitian	Corsican

Finally the main results per language are gathered in the following table:

LANGUAGE	SPOKEN BY THE WHOLE COMMUNITY	INTERNET PRESENCE	CHARACTERISTICS
Basque Corsican	Not much	Strong dynamics	Balanced deployment including local governments
Breton Franco-Provençal Occitan	Not much	Good dynamics	Citizen impulse, but needs more push from local governments
Frankish Oïl languages	Not much	Difficulties	Low interest from academia and local governments (except Gallo)
Creoles, Kanak, Futunan, Mayotte languages, Tahitian and Wallisian	Quite a lot	Weak (except Tahitian)	Pulled by academia
Alsatian Catalan	Relatively	Good	Balanced deployment including local governments

The treatment of all statistics given rise to the following categorization of languages studied:

A) Languages not much spoken within the region but with strong momentum on the Internet, with a homogeneous multi-stakeholder deployment including local government: this is the case of Corsican and Basque. In this group, Corsican is distinguished by a strong support of local authorities and strong citizen involvement, Basque is at the other side with more involvement from non profit and private sectors.

B) Languages spoken a little within the region with strong momentum on the Internet driven by citizen and not necessarily a strong local government support: Occitan, Breton and Franco-Provençal. Within this category there is a strong presence of private sector for Breton and weak and almost nonexistent

for Occitan and Franco-Provençal. In contrast, there is a high degree of citizen involvement for Occitan and still higher for Franco-Provençal.

C) Languages spoken a little within the region that are not very dynamic on the Internet: Oil languages and Frankish. Without local government support and with low interest from the academia they are supported mainly by nonprofit organizations and voluntaries.

D) Languages well spoken within the region, with low own presence on the Internet: **Creoles, Kanak languages, Futunan, languages of Mayotte, Tahitian and Wallisian**. Maintained by academia (notably outside the community) with low support from local governments. Private sector, specially from the sphere of tourism, has a strong involvement for Tahitian and Creoles, average for Mayotte, weak for Kanak languages and null for Wallisian.

E) Relatively spoken languages within the region with presence on the Internet: Alsatian and Catalan. They both enjoy a balanced deployment of stakeholders, although with very different realities. While for Alsatian, most references are proper to the community, they are alien to the Catalan (coming from Spanish Catalonia). Although local government and academic support are present and a citizen momentum arises, it may be that the activity carried out by Catalonia is largely sufficient to cover the needs of the Catalan speakers from the North and inhibit other local initiatives.

Conclusion

These are the first promising results for a field that has not yet been systematically explored. A later stage would be desirable to create an open clearinghouse where players can record their work and which is organized to promote dialogue between actors from different languages of France.

This approach should be applicable to other countries that are also experiencing a wide variety of "minority" languages such as Germany Spain, Italy or Russia.

General Conclusion

Both methodological approaches proposed in a field characterized by a prolonged crisis allow to partially overcome the existing gaps in information about the presence of languages on the Internet and make an original contribution to the subject. There is a reasonable likelihood that they can be adapted without much change to other languages than French and the languages of France.

Web References

1. Observatório da Língua Portuguesa: http://observatorio-lp.sapo.pt/pt.

2. Union Latine, "Étude langues et cyberespace": http://dtil.unilat.org/LI/2007/index_fr.htm.

3. FUNREDES, Observatoire des langues et cultures dans l'Internet: http://funredes.org/lc.

4. Summer Institute of Linguistics, "Ethnologue: Languages of the World": http://www.sil.org/ethnologue.

5. CIRAL, Centre International de Recherche en Aménagement Linguistique de l'Université de Laval, Québec: http://www.ciral.ulaval.ca.

6. ALIS/ISOC, Palmarés de la Toile (1997) http://alis.isoc.org/palmares.html.

References

1. Albuquerque, A., Esperança, J. P. (2010). El valor económico del portugués: lengua de conocimiento con influencia global. http://www.realinstitutoelcano.org/wps/portal/rielcano/contenido?WCM_GLOBAL_CONTEXT=/elcano/elcano_es/zonas_es/lengua+y+cultura/ari127-2010.

2. Barbara, W. (2000). What global language, *The Atlantic Monthly* 286: 5, pp. 52-66.

3. Calvet, L. J. (2002). Le marché aux langues. Paris. Plon.

4. Calvet, L. J. (2012). Poids des langues (Baromètre Calvet). http://wikilf.culture.fr/barometre2012/.

5. ConseilsMarketing (2010). Etude sur les blogueurs francophones. http://www.conseilsmarketing.com/e-marketing/resulats-de-letude-sur-les-blogueurs-francophones

6. Crystal, D. (1997). English as a global language, Cambridge University Press

7. Crystal, D. (2000). Language Death, Cambridge University Press.

8. Crystal, D. (2006). Language and the Internet, 2nd Edition, Cambridge University Press, (ISBN-13: 9780521868594 | ISBN-10: 0521868599).

9. DGLFF (2011). Fiche repères: Le numérique au service de la langue française et des langues de France. http://www.dglflf.culture.gouv.fr/publications/Reperes12_Numerique.pdf.

10. DGLFF (2013). Rapport au Parlement sur l'emploi de la langue française, 2013. http://www.dglflf.culture.gouv.fr/publications/Reference13_Synthese%20du%20Rapport%20au%20Parlement%20sur%20l%27emploi%20de%20la%20langue%20fran%C3%A7aise.pdf.

11. Diki-Kidiri, M. (2007). Comment assurer la présence d'une langue dans le cyberspace? UNESCO-Union latine. http://unesdoc.unesco.org/images/0014/001497/149786f.pdf.

12. European Commission (2011). Flash Eurobarometer: User Preferences On Line. http://ec.europa.eu/public_opinion/flash/fl_313_en.pdf.

13. European Union (2012). Special Eurobarometer 386 Europeans and their Languages Report. http://ec.europa.eu/public_opinion/archives/ebs/ebs_386_en.pdf.

14. Franchini, P. (2013). Le français sur l'Internet, réalisé pour la Sous-direction de la diversité linguistique et du français du Ministère français des Affaires étrangères.

15. Gomes, D., Silva, M. J. (2005). Characterizing a National Community Web. http://xldb.fc.ul.pt/daniel/gomesCharacterizing.pdf.

16. Graddol, D. (1997). The Future of English? British Council. http://www.britishcouncil.org/learning-elt-future.pdf.

17. Graddol, D. (2006). English Next, British Council. http://www.britishcouncil.org/learning-research-english-next.pdf.

18. Grefenstette, G. & Nioche, J. (2001). Estimation of English and non-English Language use on the WWW. Technical report from Xerox Corporation Center Europe. http://arxiv.org/ftp/cs/papers/0006/0006032.pdf.

19. ITU (2010). World Telecommunication/ICT Development Report 2010 – Monitoring the WSIS Targets. A mid-term review, Target 9 (content), pp. 175-192, 2010. http://www.itu.int/pub/D-IND-WTDR-2010.

20. MAAYA (2012). Net.Lang : Réussir le cyberespace multilingue, C&F Edition. http://net-lang.net/

21. Maurais, J., Morris, M. (eds.) (2003). *Languages in a globalizing world,* Cambridge University Press.

22. Millan, J. A. (2005). La lengua en el medio digital: un reto político. http://jamillan.com/lenmedi.htm.

23. OIF (2001). *Trois espaces linguistiques face aux défis de la mondialisation*. Paris, Organisation Internationale de la Francophonie.

24. OIF (2010). La langue française dans le monde, édition 2010. Organisation Internationale de la Francophonie, Nathan.

25. Paolillo, J., Pimienta, D., Prado, D. et al. (2005). Mesurer la diversité linguistique dans l'Internet, UNESCO. http://unesdoc.unesco.org/images/0014/001421/142186f.pdf.

26. Pereira, J. P. (2011). O Google quer um mundo em que todos se entendam. http://publico.pt/Tecnologia/o-google-quer-um-mundo-em-que-todos-se-entendam_1489868?all=1.

27. Pierre, J. (2007). La langue au cœur du numérique, les enjeux culturels des technologies de la langue, DGLFLF.

28. Pimienta D. (2001). Quel espace reste-t-il dans l'Internet, hors la langue anglaise et la culture "made in USA" ?, dans " Nord et Sud numériques", Les Cahiers du Numériques, Vol 2, No 3/4 Hermès, Numéro spécial sur la fracture numérique.

29. Pimienta, D. (2011). Chapter 4, Linguistic Content, in: *Making The Web More Effective For Supporting Economic And Social Development*, Word Wide Web Foundation.

30. Pimienta, D., Prado, D., Blanco, A. (2009). Douze ans de mesure de la diversité linguistique dans l'Internet: bilan et perspectives, UNESCO. http://unesdoc.unesco.org/images/0018/001870/187016f.pdf.

31. Prado, D., Pimienta, D., Lemoulinier, A. (2010). Diversité linguistique et cyberespace : état de l'art, enjeux et opportunités, Cosmopolis. http://agora.qc.ca/cosmopolis.nsf/Articles/no2010_1_Diversite_linguistique_et_cyberespace___etat_de_l?OpenDocument.

32. Suzuki, I., Mikami, Y. et al. (2002). A Language and Character Set Determination Method based on N-gram Statistics, ACM Trans. on Asian Language Information Processing, Vol 1 N3, pp. 270-279.

33. Union Latine (2010). Présence, poids et valeur des langues romanes dans la société de la connaissance, actes de la journée d'étude du 30 avril 2010 sous la direction de Daniel Prado.

Tjeerd DE GRAAF

Senior Research Fellow, Mercator Centre,
Fryske Academy
(Leeuwarden, Netherlands)

The Frisian Language and Its Presence in Cyberspace

1. Introduction

Frisian is a Germanic language closely related to English. It is spoken in Northwestern Europe, with its most important branch in the province of Friesland, in the present-day Netherlands. This variety is referred to as West Frisian in order to distinguish it from other branches in Germany (which are referred to, respectively, as North Frisian and East Frisian). West Frisian, East Frisian and North Frisian are not mutually intelligible. During the Middle Ages, Friesland was monolingual and autonomous. Old Frisian was the official language of government and many legal documents have survived from this period. From the 16th century, however, Dutch was used as the official language of the Netherlands in the halls of government, the judiciary, in education and in religion. Frisian virtually ceased being used in written form until a revival occurred at the end of the 19th century, as a result of which the language has gradually re-entered more domains.

Frisian currently enjoys official status in the Netherlands as the second language of the state and in recent decades has acquired in Friesland a modest place alongside Dutch in government, judiciary and education. Today, Friesland has some 650,000 inhabitants, half of whom are L1 speakers of Frisian, but nearly all have some understanding of the language. Thanks to the presence of Frisian in the educational system, significant numbers also have reading and writing skills, although since this provision only dates from after the Second World War, many of the elder generation, in particular, still prefer to use Dutch.

In the past language use in Friesland could be characterized as a situation of stable diglossia (Frisian (L) used in rural areas and in informal domains, and Dutch (H) in urban areas and in formal domains). During the 20th century, Dutch also gained a foothold in many L domains, primarily as a result of migration and mixed marriages. In this way, use of Frisian changed into a sort of informal (and receptive) polylingualism. Indeed, general attitudes to Frisian have become more positive, and it has become acceptable to use it in more and more domains (radio, newspapers, social media and so on).

172

The work of the Fryske Akademy (Frisian Academy) and the Mercator European Research Centre on Multilingualism and Language Learning is devoted to the study of minority languages in Europe. The Fryske Akademy focuses mainly on the history, literature and culture of the West Frisian language. This chapter considers how new technologies are used to preserve Frisian and the way in which this changes its use.

2. Frisian in Education

The role of Frisian in primary education dates back to 1907, when the provincial government offered a grant to support Frisian lessons after regular school hours. Frisian was then taught as an extra-curricular subject. Legislative provisions for Frisian only began in 1937 with amendments to the Education Act of 1920. However, Frisian was not used as an official medium of instruction. In 1950 nine primary schools began to experiment with bilingual education and in 1955 these schools became officially recognized. Frisian became an optional subject throughout the primary school and the use of Frisian as a medium of instruction was allowed in the lower grades. By 1959, the number of bilingual schools had risen to 47.

Since 1980, Frisian has been taught in all of Friesland's primary schools, where it is also used to varying degrees as a teaching medium, alongside Dutch. There is no provision for primary education entirely through Frisian, although some preschool groups are conducted exclusively in this language. At secondary level it is also possible to use Frisian as a teaching medium for some subjects, but this is infrequently done. In the early 1980s, the subject was offered by a quarter of all secondary schools on an optional basis, although only some 5% of all pupils availed themselves of this opportunity. Since 1993, Frisian has been obligatory during the first two years of secondary education.

Special projects have been initiated in the field of trilingual education. For example, Frisian, Dutch and English are currently all used as mediums of instruction at 50 Friesland's primary schools. The Fryske Akademy co-ordinates these projects and evaluates their results.

2.1. New Technologies in Education

For pre-school education, the *Tomke* project (http://www.tomke.nl) was started in 1996. Tomke is a Frisian-speaking cartoon figure, popular with young children (typically aged between two and five) who was created with the objective of promoting multilingualism. The *Tomke* project consists of books, a magazine, films and some franchise merchandise. Initially, the Tomke films were broadcast only on the regional television channel *Omrop Fryslân* and

subsequently published on DVD. However, as the films are now also shared on You Tube, it is possible for teachers and parents to show them at any time. This has facilitated a much more intensive use of Tomke films and has allowed Frisian to enter the living room of practically any children's family in Friesland.

New technologies have made Frisian education much more attractive for children and much more user-friendly for teachers. The new teaching method, *Studio F*, is currently used by over 80% of primary schools in Friesland. Since February 2013, digital material has been available in the classroom via digiboard, personal computer or tablet. The *Studio F* website (http://www. studiof.nl) gives teachers and schoolchildren access to video and audio streams, interactive games and teaching materials. The interactivity of the educational material is very attractive to children. A similar teaching method, *Freemwurk* (http://www.freemwurk.nl) is proving popular in the secondary school, with some 2,500 individual accounts created annually.

In some schools, distance learning is used and this method is particular useful when the groups of pupils are too small to finance separate Frisian language teachers. Some Frisian language teachers include Twitter activities in their classes, challenging their pupils to tweet in Frisian and to correct wrongly spelled Frisian tweets received from their peers.

3. The Frisian Media Landscape

3.1. Print Media

There is a relatively sizable literary production in Frisian, with some 100 volumes being published annually. No daily or weekly Frisian-medium newspapers exist. Frisian-medium monthly journals, such as *De Moanne* have limited circulation.

3.2. Performance Media

Friesland has one professional Frisian-language theatre company, which is very popular. Most towns and villages also have an amateur Frisian-medium theatre company. Approximately twenty CDs of popular Frisian music are released every year.

3.3. Broadcast Media

Since 1994, the regional television channel *Omrop Fryslân* has broadcast one hour of regional television per day and a total of some thirty hours of Frisian-medium television is broadcast annually all over the Netherlands

(on Sundays). *Omrop Fryslân* also provides more than eighty hours of Frisian radio broadcasting per week and some twenty minutes per week for school programmes (radio and television). *Omrop Fryslân* has a website which is visited some 700,000 times per month. It has also developed four smartphone and tablet applications which, between them, have been downloaded more than 70,000 times. *Omrop Fryslân*'s Twitter feed is followed by more than 16,000 individuals and organizations and its Facebook page has received some 4,000 "likes". Practically all this communication is in Frisian. This "media-mix" of television, radio and Internet provision has proved extremely successful.

4. Afûk and the Promotion of Frisian

The *Algemiene Fryske Ûnderrjocht Kommisje* (hereafter, Afûk (http://www.afuk. nl)) is a cultural institution in Leeuwarden which aims to promote knowledge of Friesland and the use of Frisian via the use of traditional and new media. Its editing house produces numerous Frisian-medium books, with a particular focus on educational material, and children's books, and the Frisian monthly cultural journal *De Moanne* (http://www.demoanne.nl). Afûk also organizes language courses for both native speakers and learners of Frisian and houses a special translation service, *stipepunt Frysk* where texts are translated from and into Frisian.

4.1. Afûk and New Technologies

Alongside these traditional methods, Afûk exploits new technologies. Their Twitter account (@praatmarfrysk) and Facebook page boast some 6,000 users apiece. Every year, on the third Thursday of April, Afûk organizes the Frisian Twitterday. On the 2013 Twitterday, almost 10,000 tweets were sent in Frisian to twenty-five countries as far away as the USA and Australia. The tweets were seen by over six million people. The enthusiasm of the *Praat mar Frysk* campaign motivates many to tweet in Frisian, once a year or preferably more often. The access to new media has made this campaign much more alive than it would be without. A success story any minority language can learn from.

Afûk also provides an online learning facility *eduFrysk*. This is a good example of how new technology can open up a wide range of new possibilities in language learning and teaching. Since 2010, well over 4,500 persons have applied for an account. Students with different levels of proficiency are catered for and, through its careful selection of texts, music and songs, the programme combines language learning with learning about Frisian culture. The facility also incorporates podcasts and games, which are especially appreciated by younger users. Other features include personalized profiles and virtual

communities, which enable users to chat with each other and use the language in a friendly and informal way. Specialized learning packages are developed for particular target groups, such as people working in law or medicine. Students currently following a Frisian language course and also those who have never taken a course before can all make use of *eduFryske*. Emigration from Friesland to countries such as Canada and New Zealand has led to children finding that their grandparents are speaking a language they do not understand. *eduFryske* creates an accessible way for people with Frisian roots to learn more about where their (grand)parents came from, and about the Frisian language.

Finally, Afûk provides an online dictionary, popularly known as the "*wat wurd it*", which translates words from Dutch into Frisian and vice versa. It is available on the different websites from Afûk. Afûk daily promotes a different Frisian word through *wurdboek* and various media. The use of New Media has made that Afûk has come very close to the Frisians and can approach them practically at any time.

Another success story is the introduction and the use of Wikipedia. The Frisian version of this multilingual encyclopaedia, which started in the beginning of this century now has nearly 30,000 sites and a growing number of users. Therefore, the use of new technologies has made Frisian highly accessible to people both within and outside Friesland.

5. The Fryske Akademy

The main authoritative source on the Frisian language is the Fryske Akademy. It was founded in 1938 with the aim of maintaining an academic focus on Frisian, the Frisian people and the Frisian culture. Today, it houses departments of History, Linguistics and Social Sciences.

5.1. The Department of Linguistics and New Technologies

The Department of Linguistics conducts linguistic research on all periods of Frisian. Currently, special projects are being undertaken on the phonology and grammar of Frisian and on the linguistic characteristics of Frisian spoken in urban and rural environments. The Akademy makes extensive use of new technologies. For example, it has compiled several language corpora, such as the New Frisian language corpus (25 million words), which is a digital collection of Frisian books, scientific magazines and newspaper articles. The texts in this corpus provide a tool for keeping scientific research on Frisian culture up-to-date. The corpus will eventually become freely accessible via the Internet.

The *Dictionary of the Frisian Language (Wurdboek fan de Fryske Taal* (WFT) is the product of one of the most important projects of the Fryske Akademy: the WFT-project, which collected the vocabulary of Modern Frisian (Frisian since 1800) and has been published in book form annually between 1984 and 2011. The collection was completed in 2011 and the online version (http://gtb.inl. nl)) is freely accessible via the Internet from anywhere in the world.

Other results of the lexicographic work of the Fryske Akademy are a Frisian-English, a Frisian-Frisian dictionary and dictionaries with special terminology such as the one for legal matters. The Internet has facilitated an intensified cooperation with other researchers of minority languages such as the exchange of research papers and comparison of results. The Linguistic Department has made a large contribution to the preservation of the Frisian language. Firstly, by developing the dictionaries and later on, by digitising them and developing new (online) digital applications.

Since 2011, the department has been developing the Frisian language *Taalweb*, consisting of a new online spell checker, a machine translation programme (*Oersetter*) and a dictionary portal. The whole idea behind Taalweb is to encourage people to use the Frisian language in everyday work contexts by offering user-friendly applications and including many practical examples in translations/spelling suggestions.

The Frisian Language Desk also forms part of the Linguistics Department. This service, which can also be consulted via email, is available to answer questions about spelling, phrasing or terminology and can give advice concerning the composition of Frisian texts. It also specializes in translating technical texts into Frisian such as notarial acts, and other official and technical documents. Information can be obtained about place names in Friesland and abroad, computer terminology, inland shipping and so forth.

5.2. The Department of Social Sciences and New Technologies

The Department of Social Science studies the Frisian society. The central theme of multilingualism represents a point of departure for its many projects, which include:

1. Multilingualism and minority languages

    a. A regular survey of language use in Friesland.

    b. The Frisian language abroad: the language of emigrants.

    c. Technological developments in language learning.

    d. The availability of online materials for language learning.

e. The cognitive effects of multilingualism on children.

f. Regional variation in spoken Frisian.

2. Educational research

The Department's work on multilingual education supports and evaluates education policy making, with a particular focus on the following areas:

a. The evaluation of the provincial education policy 2007–2014.

b. Language acquisition and development in young children.

c. Trilingual schools.

d. Technological developments in education.

Part of these activities take place within the framework of the Mercator European Research Centre on Multilingualism and Language Learning (http://www.mercator-research.eu), which addresses the growing interest in multilingualism and the increasing need for linguistic communities to exchange experiences and to co-operate within a European context. The Department of Social Sciences makes use of new technologies in almost all aspects of their work, using online questionnaires and social media such as Twitter, Facebook and LinkedIn.

5.3. The Department of History and New Technologies

The Department of History studies the history, literature and toponymy of Friesland, focusing primarily on historical resources. New technologies have had a big impact on its work. Collections have been digitized and are freely accessible via the Internet. One example for this is provided by the website http://www.hisgis.nl of the project HISGIS, which stands for Historical Geographic Information System. This is a digital software package, which makes it possible to elaborate geographic and historical information: Initially, the oldest cadastral maps (dating from 1832) of Friesland have been digitized and they can be linked to later versions, texts and illustrations, which in various ways can be related to each other. On the website anyone can search through historical geography and ownership maps. The Fryske Akademy is gradually completing this website with maps from other regions in the Netherlands.

5.4. The Mercator European Research Centre

The Mercator European Research Centre on Multilingualism and Language Learning is an important part of the Fryske Akademy, which addresses the growing interest in multilingualism and the increasing need of language communities to

exchange experiences and to cooperate in a European context. It gathers and mobilises expertise in the field of language learning at school, at home and through cultural participation in favour of linguistic diversity of Europe.

For all Mercator projects, Friesland is used as a living example of a bilingual laboratory. For Frisian the Centre has developed a regional dossier which presents an up-to-date description of the position of a minority language at all levels in the educational system. The structure of this dossier has consequently been used for more than forty minority languages in other EU member states. In this way the dossiers can also be used for comparative research. In 2012, the Regional Dossiers were downloaded more than 12,000 times. More information about other activities of the Mercator European Research Centre can be found on the website http://www.mercator-research.eu.

Within the Fryske Akademy, the Mercator Research Centre also takes the lead in researching the influence of new media on minority languages. Recently, Mercator has started a research on language use and social media. The research focuses on the influence of social media on language use. Firstly, the research will analyse the language use of Frisian adolescents on social media. A study of 6,000 tweets of fifty persons in this age group has just been finished. On a regular day, 13% of the tweets are in Frisian compared to 65% in Dutch. When tweets are directed to one or more addressees (starting with @) the share of Frisian messages doubles to a quarter. In this research group (twenty-four males against twenty-six females), Frisian males tweet more in Frisian than their female counterparts. At April 18th 2013, the campaign to promote the use of the Frisian language (Praat Mar Frysk) organised a Frisian Twitterday. On this day, the Frisian language is used much more by the research group. 53% of the messages are then in Frisian, compared to 29% in Dutch. To validate these results and to get an insight into language use in different contexts, demographic background data and other variables, the research will be continued with a large scale online questionnaire. The questionnaire will be both spread through social media and through secondary education.

The collected Frisian tweets are also being analysed linguistically. The input of this analysis will, among others, be used to further optimise the new spell checker that is being worked on by the Fryske Akademy. An example that can already be named now is the phonetic spelling that has been found in the analysis of the Frisian tweets. This phonetic (wrong) spelling will be included in the new spell checker. This way a large range of suggestions based on the current day spelt words can be added to the spelling checker, thus making it even more practical. Another outcome of the analysis is the regular use of code-switching: Dutch and Frisian words and characters are often mixed within one message.

As a critical note many researchers are questioning the value of social media and are concerned about the quality of the language used through social media. Social media often put limits to the physical possibilities of the user, e.g. text messaging on small mobile phones and tiny screens, or by the limits of the software, e.g. 140 characters with Twitter. For that reason the young generation feels the need to develop some kind of a "turbo" language where words are often replaced by symbols or shortened to one or two characters.

6. Concluding Remarks

Within Europe, awareness is growing of the value of linguistic diversity, the need to speak different languages and the importance of safeguarding endangered languages. The case of Frisian shows that new technologies can play an increasingly important role in the latter area. The advantage of social media is that they can strengthen the informal written use of minority languages such as Frisian amongst the young people and reinforce the sense of belonging to a minority language group. Only time will tell whether these new technologies will help save the Frisian language but, so far, the signs are positive.

Further information can be found on the following websites about Frisian and other minority languages. Part of this report is based on texts from these websites.

- Website of the Frisian Academy: http://www.fryske-akademy.nl.

- General information about Frisian and its relation to other languages: http://www.languages-on-the-web.com/links/link-frisian.htm.

- Web site by the Frisian-American author Pieter Tiersma: http://www.languageandlaw.org/FRISIAN/FRISIAN.HTM.

- Cultural institution for education in Frisian: http://www.afuk.nl.

- On the first interactive book for children: http://www.berneboek.nl.

- Web site about the use of Frisian in daily life and how to stimulate this: http://www.praatmarfrysk.nl.

- Homepage of the Mercator Research Centre: http://www.mercator-research.eu. (The site contains the series of regional dossiers, the network of schools, a database with organizations and bibliography and many rated links to minority languages.)

- The Mercator European Network of Language Diversity Centres and portal for the partners of the network: http://www.mercator-network.eu/

- Website of the Network of Schools: http://www.networkofschools. org. (The Network of Schools, a network of around 100 schools in Europe dealing with regional or minority languages in the curriculum, is maintained by the Mercator Research Centre.)

- Eurydice network: http://www.eurydice.org. (Eurydice is the information network on education in Europe. The site provides information on all European education systems and education policies.)

- Ethnologue: http://www.ethnologue.com. (Encyclopedic reference work cataloguing the world's known living languages.)

- Information on the support for regional or minority languages by the European Union: http://europa.eu.int/comm/education/langmin. html.

- European Charter for Regional or Minority Languages (1992) and Framework Convention for the Protection of National Minorities (1995) European Treaty Series/Série des traités européens ETS 148 and 157, Strasbourg: http://conventions.coe.int/.

- Foundation for Endangered Languages: http://www.ogmios.org.

References

1. Bangma, I., van der Meer, C., Riemersma, A. (2011). *Trilingual Primary Education in Europe; some developments with regard to the provisions of trilingual primary education in minority language communities of the European Union*. Fryske Akademy publication.

2. Cenoz, J., Gorter, D. (2005). Trilingualism and minority languages in Europe. In: *International Journal of the Sociology of Language* 171: 1-5.

3. Gorter, D. (2012). *Minority languages and new technologies: solutions and threats*. Unpublished paper presented at the European Expert Seminar on Social Media and Lesser Used Languages, Fryske Akademy.

4. De Graaf, T., Tiersma, P. (1980). Some phonetic aspects of breaking in West-Frisian. In: *Phonetica* 37: 109-120.

5. European Charter for Regional or Minority Languages. *Explanatory Report*. (ETS no. 148) (1998). Strasbourg: Council of Europe.

6. Extra, G., Gorter, D. (2008). The constellation of languages in Europe: an inclusive approach. In: G.Extra and D.Gorter (eds.), *Multilingual Europe: Facts and Policies*, 3-60. Berlin: Mouton de Gruyter.

7. Gorter, D. (2005). Three Languages of Instruction in Fryslân. International Journal of the Sociology of Language 171: 57-73.

8. Jones, E. H. G., Uribe-Jongbloed, E. (2013). Social Media and Minority Languages. Convergence and the Creative Industries. Bristol: Multilingual Matters.

9. Jongbloed-Faber, L., Van der Meer, C. & Klinkenberg, E. (2013). Language use and social media of Frisian Adolescents. Unpublished raw data. Fryske Akademy.

10. Riemersma, A., de Jong, S. (2007). Frisian. The Frisian language in education in the Netherlands (4th edition). Ljouwert: Mercator Education [Regional Dossiers Series]. On-line at www.mercator-research.eu.

11. Sijens, H. & Dykstra, A. (2013). Language Web for Frisian. E-Lex 2013 conference on electronic lexicography. Talinn. (Forthcoming).

12. Tiersma, P. M. (1985, 1999). Frisian Reference Grammar. Dordrecht: Foris Publications.

13. Ytsma, J. (2007). Language use and language attitudes in Friesland. In: D. Lasagabaster and A.Huguet (eds.), Multilingualism in European Bilingual Contexts (Language use and attitudes), 144-163. Clevedon: Multilingual Matters.

Harald HAMMARSTRÖM

Researcher, Max Planck Institute for Psycholinguistics (Netherlands)
(Stockholm, Sweden)

Glottolog: A Free, Online, Comprehensive Bibliography of the World's Languages

Glottolog (http://glottolog.org) is a *bibliography* of descriptive materials on the languages of the world (this part is also known as LangDoc) and a *classification* of the world's languages. Glottolog is browsable, searchable, downloadable, continually updated and free of charge.

The Glottolog bibliography was created in response to the lack of a sufficiently comprehensive and accessible bibliography on a world-level scale. Other large bibliographies exist already but have drawbacks on one or more of the desired aspects. For example, the SIL bibliography[74] accompanying the Ethnologue [Lewis et al. 2013] has a large number of bibliographical entries on lesser-known languages, but is restricted to work produced under the Summer Institute of Linguistics (SIL) umbrella. The Bibliographie Linguistique[75] is not restricted to a certain producer but systematically fails to include MA/PhD theses as well as items from minor countries, both of which make up a significant part of the total. Also, use of the Bibliographie Linguistique is not free of charge. Worldcat[76] has an enormous collection but also lacks large classes of items from major countries. Also, there is no systematic way of singling out linguistically relevant publications nor languages. Google Books[77] may have an even bigger coverage but, similarly, when it comes to lesser-known languages, there is no systematic way of singling out linguistically relevant publications nor languages.

The philosophy of comprehensiveness of the Glottolog bibliography is as follows:

A: Include the *most extensive pieces* (MED) of documentation for every language, and,

B: Beyond that, include "as much as possible".

---

[74] *Available at http://www.ethnologue.com/bibliography.asp.*

[75] *Available at http://bibliographies.brillonline.com/browse/linguistic-bibliography.*

[76] *Available at http://www.worldcat.org.*

[77] *Available at http://books.google.com.*

This implies that for a small language with only a wordlist to its documentation reference should be in Glottolog. For a bigger language with countless articles/ books, a major dictionary/text/grammar collection should be included, but not necessarily every reference ever written about the language (but, of course, any amount of these are also welcome). In essence, only published or publicly accessible materials are considered, as opposed to ongoing unpublished work or manuscripts whose existence/access is difficult to confirm. Master's theses and PhD theses are included since they are in principle accessible from the approving institution.

Since there are over 7,000 languages in the world, the practice of actually obtaining the reference to the most extensive pieces of documentation for every language is highly non-trivial. In addition, language documentation and description is an extremely decentralized activity, carried out by missionaries, anthropologists, travellers, naturalists, amateurs, colonial officials, ethnographers and not least linguists over several hundred years. In fact, there has never in history been a systematic survey of descriptive materials on the languages of the world (although, at least, Adelung [1820] and Schmidt [1926] were in a position to produce one at times when the task was much smaller).

A legitimate question, then, is who knows all the obscure bibliographical references? The Glottolog answer is: experts on language families or areas know the bibliographical references for the corresponding families/areas. Experts write handbooks and overviews, such as outright bibliographies, comparative/ descriptive overviews, and sociolinguistically oriented overviews. Following this logic, one may go through all handbooks and overviews and collect the references to obtain a comprehensive collection (for more details see [Hammarström and Nordhoff 2011]). The task of going through all handbooks/ overviews is not necessarily a lesser amount of work because there are more handbooks/overviews than the number of languages (over 8,100, also listed in Glottolog and tagged as such). But it is more systematizable since countries, areas and families are easier to enumerate.

In addition to the strictly systematic collection, Glottolog also incorporates any existing bibliography available. A selection of the largest bibliographic databases granted by their respective compilers are listed in Table 1. A complete list with full descriptions of the source bibliographies and their provenance is available at http://glottolog.org/langdoc/langdocinformation. The total amount after removing duplicates is currently 193,407 references. This is not everything that has ever been written about any language, but the most extensive description for every language is included.

Table 1. Some existing bibliographical resources and their size, contents, annotation and the time the information was culled

	Number of references	Contents	Area	Coverage	Annotation		Date
EBALL	60,164	Everything	Africa	Full	100%	L & T	2009
HH	34,197	DD	World	85%?	100%	T	2014
Fabre	30,176	Everything	S. America	Full	100%	L	2009
SIL	18,464	Mainly DD & VP	World	70%?	100%	L & T	2009
MPIEVA	13,966	Everything	World	?	62-93%	L & T	2009
SILPNG	13,110	Mainly DD & VP	Papua	Full	100%	L & T	2004
ANLA	11,627	Mainly DD & MSS	Alaska	Full	100%	L	2012
OZBIB	10,377	Mainly DD	Australia	Full	100%	L	2010
WALS	5,633	Mainly DD	World	?	99%	L	2005

L = Language, T = Type, DD = Descriptive Data, VP = Vernacular Publications, MSS = Manuscripts

For enhanced sorting and filtering capabilities, references in Glottolog have some annotation. From the searcher's viewpoint, the more and the more detailed content-annotation the better, but from the annotators' viewpoint, more and more detailed annotation is more work, unless the annotation can be (semi-)automatized. In general, we only have access to the text of the bibliographical reference itself (author, title, year, etc.), not the actual document it refers to. Therefore, inferences depending on page counts or words that tend to occur in the title are possible, e.g., the name of the language(s) being treated often appears in the title (see below), but we cannot tell, e.g., whether there is a chapter/section on adjectives or whether numerals are included in a wordlist. As a compromise between search desiderata, annotation work and (semi-)automatizabillity, Glottolog references are annotated as to language and description type. As to language, references are tagged with ISO-639-3 language code(s)[78] which allows lookup for location, speaker numbers, etc. Ideas on annotation also for other levels (above the language level, i.e., a

---

[78] *See http://www.sil.org/iso639-3/default.asp.*

(sub)family or below the language level, i.e., a (sub)dialect) are currently being implemented (cf. [Cysouw and Good 2013]). As to description type, Glottolog references are annotated according to the hierarchy in Table 2. Roughly half of the references in Glottolog are manually annotated, often by translation of an annotation scheme used by a source bibliography, and the other half is automatically annotated based on words in the title of the reference. Essentially, a reference titled "A grammar of Tauya" can be inferred to be of the description type grammar and the language Tauya (see [Hammarström 2008, 2011] for details). Since the automatic annotation has a much higher error rate than manual annotation, it is recorded for every reference which annotation is automatic and which is manual, and this can be used for filtering. However, quality assessments requiring specialized knowledge fall outside the current scope of Glottolog.

Table 2. The typology of description types used in Glottolog

Type	Explanation
grammar	a description of most elements of the grammar ~ 150 pages and beyond
grammar sketch	a less extensive description of many elements of the grammar ~ 50 pages
dictionary	~ 75 pages and beyond
text	text material
specific feature	description of some elements of grammar (i.e., noun class system, verb morphology, etc.)
wordlist	~ a couple of hundred words
minimal	A small number of morphemes
overview	Document with meta-information about the language (i.e., where spoken, non-intelligibility to other languages, etc.)

The Glottolog website has entry points for simple searches on references, simple searches on languages and (sub)families and complex searches for reference and (sub)family at the same time. In the latter way one can, e.g., search for all grammars for African languages of the Semitic subfamily produced in 1984.

An example of a typical Glottolog view is shown in Figure 1 focussing on the language Sakha (also known as Yakut). The tree showing the classification of Sakha is shown at the top left and the geographical location on the map on the right. The tree is navigable and the references listed below provide the justification for why Sakha is (sub)classified the way shown. The classification

employs an even standard of evidence required across all families/languages/ areas of the world[79]. Since the evidence for most families can be debated to some degree, pointers to the arguments supporting each node are given in brief with references to literature along with comments if needed. The list at the bottom contains the bibliographical references tied to Sakha. The list can be filtered, sorted and downloaded in various formats.

Figure 1. *An example of a typical Glottolog view*

The entire Glottolog database, both the classification and the bibliography, can be downloaded[80] along with older versions. Versions starting from Glottolog 2.3 are long time archived with a DOI[81]. The data is also available as Linked Open (see [Forkel 2014]).

References

1. Adelung, F. (1820). *Uebersicht aller bekannten Sprachen und ihrer Dialekte*. St.Petersburg: Nic. Gretsch.

2. Cysouw, M. & Good, J. (2013). Languoid, Doculect, Glossonym: Formalizing the notion "language". *Language Documentation and Conservation* 7. 331–359.

---

[79] *For more detailed information, see: glottolog.org/glottolog/glottologinformation.*

[80] *From the downloads page http://glottolog.org/meta/downloads.*

[81] *See http://dx.doi.org/10.5281/zenodo.*

3. Forkel, R. (2014). The Cross-Linguistic Linked Data project. In: C. Chiarcos, J. P. McCrae, P. Osenova & C. Vertan (eds.), *3rd Workshop on Linked Data in Linguistics: Multilingual Knowledge Resources and Natural Language Processing*, 60–66. Reykjavik, Iceland: European Language Resources Association (ELRA).

4. Hammarström, H. (2008). Automatic annotation of bibliographical references with target language. In: *Proceedings of MMIES-2: Workshop on Multi-source, Multilingual Information Extraction and Summarization*, 57–64. ACL.

5. Hammarström, H. (2011). Automatic Annotation of Bibliographical References for Descriptive Language Materials. In: P. Forner, J. Gonzalo, J. Kekäläinen, M. Lalmas & M. de Rijke (eds.), *Proceedings of the CLEF 2011 Conference on Multilingual and Multimodal Information Access Evaluation* (LNCS 6941), 62–73. Berlin: Springer.

6. Hammarström, H. & Nordhoff, S. (2011). LangDoc: Bibliographic Infrastructure for Linguistic Typology. *Oslo Studies in Language* 3(2). 31–43.

7. Lewis, P. M., Simons, G. F. & Fennig, C. D. (2013). *Ethnologue: Languages of the World*. 17th edn. Dallas: SIL International.

8. Schmidt, W. (1926). *Die Sprachfamilien und Sprachenkreise der Erde* (Kulturgeschichtliche Bibliothek. Reihe 1, Ethnologische Bibliothek 5). Heidelberg: Carl Winter's Universitätsbuchhandlung.

Dietrich SCHÜLLER

Chair, Working Group on Information Preservation,
UNESCO Information for All Programme
(Vienna, Austria)

Magnetic Tape Apocalypse:
Safeguarding the Documents Proper of Linguistic and Cultural Diversity

"Audio and Video Documents at Risk" was the title of a presentation by the author at the Second International Conference on Linguistic and Cultural Diversity in Yakutsk, July 2011[82] .

This paper described the unique role of audiovisual records as the documents proper of linguistic and cultural diversity of human kind, and their indispensability for the documentation and study of spoken language, music, dance, rituals, and other optical or acoustical cultural phenomena. Consequently, many research disciplines have only emerged with the advent of audiovisual recording technology at the end of the 19th century. Since the 1950s, portable audio and since the 1980s video recording equipment has lead to a mushrooming of audiovisual collections world-wide which now form the basis of our present knowledge in their respective disciplines. In contrast to conventional text documents, audiovisual documents are subjected to a bundle of specific risks: they are unstable, and, as machine readable documents in need of specific replay equipment. This situation ultimately makes conventional conservation of original documents impossible, which has lead already around 1990 to a change of the preservation paradigm: to concentrate on *content* preservation by retrieving the contents from their original carriers, by converting them to digital files, and to preserve these files by "eternal" digital and thus lossless migration from one preservation platform to the next. By adopting this paradigm, gradually audio, video, and recently also film preservation became part of the IT world. This transition process, colloquially termed digitisation, is logistically and financially demanding. Broadcast and national archives of wealthy countries have already started, or even finished this process, while collections in developing countries are lagging behind because of notorious lack of funds. Specifically threatened – in all parts of the

---

[82] *Audio and Video Documents at Risk: Safeguarding the Documents Proper of Linguistic Diversity and Orally Transmitted Cultures. In: Linguistic and Cultural Diversity in Cyberspace. Proceedings of the 2nd International Conference, Yakutsk, 12–14 July 2011. Moscow 2012, 151–157.*

world – are research documents, as most of them are held outside archival custody in the narrower sense, which results in unawareness of the risks and of the urgency to act.

As compared to 2011, the situation in 2013/2014 looks even more dramatic: the availability of audio and video equipment in operable condition is shrinking with frightening speed. In 2011, the remaining time window was quoted to be "10-15 years", but this needs to be revised, specifically for magnetic tape replay equipment. The most recent trigger for enhanced alert is the last production run of replay heads for Studer A 807, the world's most popular and widespread modern magnetic audio tape replay machine used for the digitisation of analogue magnetic audio recordings. The replay head is the heart of signal extraction process, the part that converts the magnetic information on the tape into the electric signal to be digitised. Replay heads are high precision spare parts, their quality is highly dependent on specialised production machines and – most important – on the experience of the operating personnel. Upon initiative of the Technical Committee of the International Association of Sound and Audiovisual Archives (IASA), a collective order for 600 replay heads was placed for a last production run before the company is going out of business. The life time of a head is 2-3000 hours, and it is unlikely that additional precision heads can ever be produced again once the recent supply has been used up.

Replay head of a Studer A 807 audio tape replay machine

Audio and video replay heads are of central importance for magnetic tape replay devices, however they are not the only spare parts that become unavailable: sensors, motors, and breaks are spare parts of high specialisation which can

190

hardly be produced individually, and failing integrated circuits ("chips") are irreplaceable once production has been ceased. Analogue audio tape machines need reference tapes for their alignment, for which there is only one producer left. There are also more trivial objects like belts, pulleys and pinch rollers, and accessories like leader and splicing tape, spools and cassette shells that are essential for the replay of audio and video tapes, but increasingly unavailable.

The pessimistic spare part situation is aggravated by fading professional services and skills. Producers discontinue maintenance of their equipment of meanwhile obsolete formats, and highly specialised engineers reach retirement age. Today, responsibility for professional maintenance is moving to audiovisual archives, but only few of them are able to continue such services to the extent and level formerly performed by the producers of the equipment.

Less dramatic is the situation for the replay of mechanical audio carriers, specifically microgroove discs, so-called LPs or "vinyls". After CDs had taken over in the 1980s, there is presently a "vinyl revival" which at the moment lifts pressure from mechanical disc replay. But there is no realistic hope for a similar movement in magnetic tape replay.

In summarising the situation, we have to realistically assume that within few years even the most popular and wide-spread magnetic audio and video tape formats cannot be replayed anymore, a situation which is already true for many video tape formats like Video 2000, Betamax or MII which at their times have not reached sufficient market acceptance. This may lead to a significant amount of magnetic tape collections which, despite their physical integrity, will be lost, because of the lack of replay equipment. This development will reach beyond the professional world: Compact Cassettes and MiniDV tapes, massively produced and used for private purposes, will also be affected.

In order to prevent this loss – or at least keep it as small as possible – the Information Preservation Working Group of the Information for All Programme (IFAP) suggested UNESCO to embark on an awareness raising campaign. This campaign entitled *Magnetic Tape Apocalypse* will inform governments and stakeholders of the threat to documents stored on magnetic tapes. IFAP and Memory of the World communities, archival NGOs and academic societies will be involved to assess the qualitative and quantitative dimensions by a short questionnaire on the UNESCO website. On the basis of this information a concrete Plan of Action will be developed.

The campaign will start in autumn 2014, feed back is expected by the end of 2014, which will result in a Plan of Action in early 2015. This plan will go beyond awareness raising and will highly depend upon the response by Member States and concerned academic/cultural institutions and organisations.

In summarising it may be reminded that one of the key missions of UNESCO is the protection and promotion of linguistic and cultural diversity of humankind. Written documents are unable to adequately represent spoken languages, dialects, and orally transmitted cultural phenomena like music, dance, and rituals. Therefore, audiovisual documents are the documents proper of linguistic and cultural diversity, and, consequently, their preservation is closely related to this key mission of UNESCO.

Failure to preserve audio and video documents would be an unprecedented act of passive destruction of primary source materials, thus undermining fundamental research principles. Without these primary sources, the validation of our present day's knowledge would be equally impossible as the new interpretation of these sources in the light of evolving scholarly and cultural interests.

Their loss, eventually, would considerably diminish the resources for linguistic and cultural diversity in cyberspace

References

1. IASA Technical Committee Publications (http://www.iasa-web.org/iasa-publications):

    - The Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy, edited by Dietrich Schüller. (= IASA Technical Committee – Standards, Recommended Practices and Strategies, IASA TC-03), Version 3, 2005. Also available in German, French, Swedish, Spanish, Russian, Italian and Chinese.

    - Guidelines on the Production and Preservation of Digital Audio Objects, edited by Kevin Bradley. (= IASA Technical Committee – Standards, Recommended Practices and Strategies, IASA TC-04). Second edition, 2009.

    - Handling and Storage of Audio and Video Carriers, edited by Albrecht Häfner and Dietrich Schüller. (= IASA Technical Committee – Standards, Recommended Practices and Strategies, IASA TC-05), 2014.

2. Schüller, D. (2008). Audiovisual research collections and their preservation. http://www.tape-online.net/docs/audiovisual_research_collections.pdf.

3. Schüller, D. (2012). Challenges for the Preservation of Audiovisual Documents. A General Overview. In: L. Duranti and E. Shaffer (eds.), Proceedings of the Conference "The Memory of the World in the Digital Age", 26–28 September 2012, Vancouver, Canada. 863–869. http://www.ciscra.org/docs/UNESCO_MOW2012_Proceedings_FINAL_ENG_Compressed.pdf.

Adolf KNOLL

Secretary for Science, Research and International Cooperation,
National Library of the Czech Republic
(Prague, Czech Republic)

Manuscriptorium.
International Aggregation of Multilingual Content within Digital Library

Digitization of library collections started at the National Library of the Czech Republic quite a while ago. The impetus was given by several projects, the most significant of which were the earliest ones which aimed to support the Memory of the World UNESCO Programme in 1992 and 1993. Several products which appeared as a result of this activity showed that digital technologies, even at an early stage of their development, could contribute to protecting rare manuscripts by excluding direct physical contact with them and slowing down their damage which may occur in the process of their study.

It was this understanding that led us to establishing the first Manuscript Digitization Centre within the National Library in 1995–1996. Since then, digitization of manuscripts has been part and parcel of our routine library activity. In 2000, thanks to the donation programmes of the Ministry of Culture of the Czech Republic, we managed to launch several digital data production subprogrammes among which were the ones for manuscript and old rare printed books digitization. In addition, we launched a retrospective catalogue conversion subprogramme, a periodicals protection and digitization subprogramme and some others. As a consequence, with several annual projects taken as a basis, digitization of manuscripts was launched by other research and public libraries, museums, archives, monasteries, and, also, the libraries of Czech castles. Thus, the project launched by the National Library of the Czech Republic has acquired a national status which implies introduction of unified technological standards.

The gradually growing interest of researchers to the online access to digital copies of manuscripts and rare books gave us an impetus to launch Manuscriptorium, a digital library of manuscripts. It was in 2003. The documents which were digitized within the above national programme made the bulk of the library's collection, but there are also significant catalogues of manuscripts. We have detected interest in enriching the library by information about manuscripts and rare documents which are related to our cultural tradition. This interest

has been demonstrated by not only institutions of our country but some foreign libraries as well. It is worth noting that the borders of the countries in the Central Europe were subject to frequent changes, and many countries made part of the Holy Roman Empire on whose territory one may observe the development of similar approaches to the creation of manuscripts and a broad cultural exchange. The territory of the Holy Roman Empire embraced, at different times, that of contemporary Germany, Austria, Czechia, Slovenia, Switzerland and Luxemburg, Northern Italy and Western Poland Silesia). Moreover, Western Christianity served as a media for cultural similarity, which allows us to speak of the Western European manuscript tradition in opposition to Slavic (Slavic manuscripts in the Cyrillic script), Greek and Jewish traditions. In the course of the centuries, lots of these manuscripts have changed their location due to warfare and other events, making it especially important to unite them at least virtually within a unified digital library. Additionally, European libraries may hold non-European manuscripts and old books, like it is done by the National Library of the Czech Republic which holds interesting Arabic, Persian, Ottoman and Indian manuscripts.

These are the reasons for which some foreign libraries started joining Manuscriptorium. The first among them was the Wroclaw University Library in Poland. Aggregation from abroad has been developing within the ENRICH European project. It was led by our library in 2007–2009 and included 18 partners from several European countries. Manuscriptorium has grown since then thanks to some subsequent European projects within the CULTURE programme and bilateral initiatives.

Metadata Agenda

In1995, we decided to introduce daily digitization of manuscripts. It immediately became clear that we needed to be maximally free from the influence of specific software on digital content. At that time, there were few approaches in the world that could be considered reliable and comprehensive enough for our task. So, we decided to take our own route and design our own SGML-based approach. That was how a new language appeared. We called it DOBM and used until 2002. Any approach of ours had to contain a document description and reflect the document structure. It had to represent each document by means of digital images which we decided to supplement with some technical information for the purpose of getting true simulation of a digital copy in a computer environment so that each manuscript could be displayed in the form it had at the time of digitization. In 1999, we produced a CD with our approach to digitization of various documents and demonstrated that it was possible to use this approach for processing audio records. Our

approach was recommended by UNESCO as a model one for the Memory of the World UNESCO Programme[83].

In 2002, we changed the platform for TEI. It became possible thanks to the development of a new approach to the electronic description of manuscripts. This approach was a product of the European MASTER project in which we also played a part. We added to the so-called master.dtd[84] some structural elements and some elements that allowed for keeping technical information about digitization and the digital images themselves. The thus extended standard was called masterx.dtd[85]. It was used virtually as a national manuscript digitization standard until 2009 when it was replaced by enrich.dtd[86], a standard developed within the ENRICH project on the TEI P5 platform.

Unlike previous standards, enrich.dtd was created as a data exchange standard or, to be more precise, as a unified platform for the aggregation of digitized manuscripts and old printed books. Therefore, enrich.dtd supports only the codification of the manuscript description and structure, including those images which serve for the structure implementation. Meeting the needs of the ENRICH partners, this standard was implemented as an internal standard of Manuscriptorium, i.e. not only as a data exchange format but also as a data aggregation format within a unified database. This became possible because, while designing this standard, we managed to meet the requirements of the two main approaches to manuscript description in the electronic environment: those of the library and researchers.

From the standpoint of the requirements to the description of manuscripts, the library approach, which is based mainly on the MARC format and its variations, is insufficient and elementary. Therefore, at the end of the 1990s, there appeared the first attempts to use the TEI language. Unfortunately, the first project, which was produced in this language (master.dtd), did not take into account some specificity of MARC formats, especially their formal approach to recording the names of physical persons, though it suggested a detailed and even hierarchical description of the manuscript content. Both approaches allowed for carrying out all necessary operations with manuscripts in two fields: their elementary recording with libraries and their scientific description by researchers. However,

---

[83] *Digitization of Rare Library Materials: Storage and Access to Data / Project Management by Adolf Knoll and Stanislav Psohlavec. Authors: Adolf Knoll, Stanislav Psohlavec, Jan Mottl, Jan Vomlel, Tomáš Mayer. Prague, National Library - Albertina icome Praha, 1999. Memoriae Mundi Series Bohemica, also online at: http://digit.nkp.cz/knihcin/digit/WWW/ENTER.HTM.*

[84] *http://www.tei-c.org/About/Archive_new/Master/Reference/oldindex.html.*

[85] *http://digit.nkp.cz/MMSB/1.1/msnkaip.xsd.*

[86] *http://projects.oucs.ox.ac.uk/ENRICH/.*

when we decided to combine these approaches within a unified database, we revealed some loss of information. Since MARC doesn't allow for a sufficiently deep and detailed description and, at the same time, flexible extension, we found a decision on the basis of the TEI platform.

In 2012–2013, in order to produce data ourselves, we worked out detailed requirements to the creation of the information batch (definition of a digital document with regard of access requirements and permanent storage within the VISK6 subprogramme)[87]. We determined the structure of the batch (i.e. what and where should be located). Undoubtedly, its essential constituent was the enrich.dtd format, which was supplemented by some other features: the rules of assigning unique file names which are part of the manuscript's digital format; the rules of recording technical data about digital images and other kinds of technical data such as ICC profiles; tool calibration data; the need in having two specific calibration matrices and recording of colorimetric features of the colors represented in them.

We had to take into account that each change in the manuscript format resulted in a complex migration of the entire digital library content and changes in the "environment", i.e. the programmes and tools which support document production, indexing and displaying, to say nothing of handling them at the user's end.

Aggregation Principle

The intention to concentrate in the digital environment all of the requested information about resources appeared right after the emergence of the Internet: the decisions how to do it may be different but they all have common features, i.e. accumulation of information from various resources and their indexing and an opportunity for conducting search in order to meet user information requirements. Search engines appeared for this purpose, like the currently available Google, portals and even digital libraries. When the latter are referred to, classification challenges turn up: for instance, portals are often called digital libraries, though they are none of the kind: Internet search engines work with all accessible Internet wealth, while portals aggregate information from collaborating resources only. In both cases, users find meta-information which, if necessary, links them with the required resource which is accessible for processing. In the cultural environment, such portal is EUROPEANA. From the standpoint of the user, who needs to spend time and efforts to access the

---

[87] *Definice digitálního dokumentu pro potřeby zpřístupnění a trvalého uložení v podprogramu VISK6 / Olga Čiperová, ŠtěpánČernohorský, TomášKlimek, TomášPsohlavec, online at: http://www.manuscriptorium.com/ sites/default/files/docs/manuscriptorium_visk6_definice.pdf.*

resource in need, the portal replaces his/her physical journey from one library to another by a virtual journey from one resource to another. The activity of librarians or collection curators is replaced by the nature of the resource, its interface specifically. In any case, libraries and interfaces may have their own specific features, and each demands a special user approach.

Intending to simplify the work of our users, we strive to aggregate not only meta-information/metadata but also information as such, i.e. data or files with texts, images, audio and video records. The simplest way to do it is by concentrating all information in one databank, one physical digital storage, but for this we must come to an agreement with the owners and transform every document we get from them to the format of our resource and download it in full into our digital storage. Such is the work of the World Digital Library of the Library of Congress. Due to the labor intensiveness of this process the Library of Congress may obtain only a small share of the digital collections of its partners.

Another solution is aggregation of those metadata which allow for not only obtaining information about the nature and location of the required document but also understanding how it is structured in the partner's digital storage whose main purpose is to meet the requirements of the partner's digital library. Each digital library has (1) a metadata database in which users can search and which contains information about the whereabouts of the data in question (in our case, these are texts or images of the manuscript pages), and (2) a digital storage of these data, from which users obtain the images or texts of the required manuscripts.

If, in the partner's storage, image files are given fixed invariable addresses, as is the fact in the vast majority of cases, these image files may be requested by various data presentation systems, for example, through various digital libraries which contain data about the location and the way these image files can be pooled within a manuscript.

Aggregation in Manuscriptorium is based on this principle. It allows users to work in real time with the data from different digital libraries within the Manuscriptorium's unified interface. You may get an impression that all data are located in one place, while Manuscriptorium travelled from one digital storage to another to harvest them. Such aggregation of data from various resources is called seamless aggregation, and it rests on the exhaustive agreement with the partners.

Manuscriptorium prefers to get the metadata required for aggregation in an automated mode via the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).It demands that the OAI profile includes a document

description and structure with references to the files out of which it is possible to represent it in the digital environment. At our end, we offer to every partner special tools for these metadata to transform/convert them from the partner's format into the unifying internal format of Manuscriptorium. In our terminology, a set of tools corresponding to a certain partner is called a connector. The advantage of our approach is that connectors are developed only once and forever, and that harvesting/collection of metadata via OAI is done on a regular basis. As a result, when a partner adds new documents/ manuscripts to his/her digital library, these documents/manuscripts are added to the Manuscriptorium's virtual collection after certain procedures (harvesting and processing for Manuscriptorium).

Undoubtedly, there are other methods of cooperation for the cases when a partner has neither an opportunity to work with OAI nor his/her own digital library. For the cases like these Manuscriptorium offers a set of online tools which can foster the production of descriptions in the Manuscriptorium's internal format, their maintenance and downloading to Manuscriptorium.

User Environment Content and Customization

Manuscriptorium contains about 331,000 descriptions, more than 25,000 of which refer to digitized documents which are accompanied by more than 600 full texts. The total number of pure full texts exceeds 2,000 because some documents are represented by texts only, without any images.

Manuscripts are the main part of the digital library, but there are also extremely rare printed editions including geographical maps. The data to Manuscriptorium have been provided by some 120 institutions, 55 of which are from the Czech Republic. Almost 70% of the fully digitized documents are provided by our foreign partners. Our library is the major provider of documents (over 3,500); other big partners in our country are the Moravian Library in Brno, the Strahov Monastery in Prague and the National Museum Library; they are followed by many organizations of various kinds, including the Kynžvart Castle Library (Königswart in Western Czechia), i.e. the library of Prince Metternich, Foreign Minister of the Austrian Empire. Among foreign partners the biggest content-providers (by the number of documents provided) are the Complutence University of Madrid (Spain), The Holy Trinity-St. Sergius Lavra (Russia), the Wroclaw University Library (Poland); National Libraries of Italy (Florence), Spain, Iceland and Romania and University Libraries of Vilnius (Lithuania), Heidelberg (Germany), Bratislava (Slovakia), Zielona Góra (Poland) and many others. In some cases, our partners administer the group of participating libraries themselves (Cologne University, Germany)

and undertake a certain national aggregation (first and foremost, the eCodices Swiss project).

As for the quantity, ranked first are West European manuscripts. However, thanks to the efforts of some of our partners, Slavic manuscripts and rare books also constitute a significant part of the collection (Sergiev Posad, National Library of Serbia, Research Library in Plovdiv, National Library of Romania and even our Slavic Library). Additionally, the digital library contains some significant manuscripts in Hebrew, Arabic and Farsi.

The digital library's versatile content brings about new challenges connected with access to documents because library descriptions are made in modern living languages of collection curators or, in case of some Latin manuscripts, in Latin. This limits sometimes the perception of the manuscript content. For instance, a Romanian curator, when reading a Czech manuscript, is able to read the text and decide that the manuscript is about a Slavic language. He/she will write this down in Romanian in the document description, but he/she is unable to determine that the manuscript was written in the Czech language of the 15th century. Further challenges appear, in the first place, when the text of the manuscript is written in a non-Latin script like Hebrew, Greek, Arabic or Cyrillic, and when the romanization (transliteration) of the title is made according to the rules which are in force in a certain organization or in a certain cultural or linguistic environment. In case of the Cyrillic script, it is especially hard to match, say, Czech/Mid-European, Romanian, English romanization and an entry without such. Moreover, a manuscript is not a standard published book, which forces us to use artificial titles of handwritten volumes. As is well-known, when a Latin manuscript includes various texts it is often called merely "Textus varii" or, when a manuscript is extremely large, it is called "Codex Gigas". As for the manuscripts in other languages, an artificial title is written simply in the cataloguing language. Thus, in Manuscriptorium, it is possible to find artificial titles of Slavic manuscripts in different languages, for example in Romanian (Molitvenic slavonesc, i.e. Slavic player book), in Russian in Czech romanization (Pritči o četyrech vremenach goda, i.e. Parable of Four Seasons) or directly in Russian in the Cyrillic script (Октоих на линейных нотах, i.e. Octoechos on Ledger Lines). Sometimes you may come across Romanian romanization of the Cyrillic script which looks as follows: Evanghelie naprestol'noe, napec. Diakom' Koresi i Manuilom, bez oznacenija measta pecatanija...

The same happens with the Arabic script. So, it is very hard to combine everything technologically. In my opinion, the way out of this situation is to link users to the content organization process by customizing user environment and publishing the results obtained in this environment.

Some time ago, we studied machine translation capabilities and came to a decision that machine translation might be used for the cataloguing/description language only. This is a modern language while the language of many documents may be at different levels of the historical evolution, and spelling rules at these levels could be different. Moreover, in the handwritten format, the spelling rules were not observed very often. With this in mind, we consider machine translation inexpedient in this context and even unworthy, because the researchers who wish to study a certain document have the required cultural and linguistic knowledge. As for the cataloguing or description language, it can be quickly translated in the Internet with the help of easily accessible tools in order to get the required information which is added to the manuscript by a cataloguer or a researcher working with the respective manuscript.

From another standpoint, it is essential to improve search capabilities with account of all disproportions, defects and possible violations of the spelling norms of certain languages which took place in the course of their development. As for the Latin script, a serious challenge is the use of diacritics. In historical texts and in the field of document processing out of their respective linguistic environment, we often come across an irregular use of diacritics: diacritic symbols may be used irregularly even within one word in various documents and representations. As a result, it may be hard to get a correct answer to a correctly put question even if the information in the database indices is precise and was entered by the authors, scientists and librarians. A lot has already been done in this respect but there is a lot more to do.

Every user can easily and freely register with Manuscriptorium and get an opportunity to create his/her own library on the basis of any content of the digital library: he/she may keep in it his/her information queries which gives him/her an opportunity to (1) enrich automatically his/her search results by new digital library acquisitions provided by all its partners, (2) create his/her own collections of documents without any geographical restrictions in a unified environment, and, (3) irrespective of certain volumes being dispersed throughout Europe, create virtual documents by selecting separate pages from any documents out of any collection and adding them to his/her new virtual book. Thanks to seamless aggregation of content, all the thus created products will behave homogeneously and leave an impression that they are in front of us in one place though they do not exist in real life. Every single user can dynamically create again and again his/her virtual documents in his/her interface within Manuscriptorium, travelling virtually throughout Europe being unaware of that fact. Users can exchange their results and, in future, they will be able to improve the content of the digital library and exchange their knowledge within it.

Manuscriptorium in a Broader Information Environment

Manuscriptorium is the European largest digital library of manuscripts. This and the fact that the library's documents originate from the collections of various foreign countries explain why Manuscriptorium is used by many international portals and professional services including libraries and cultural institutions and organizations: EUROPEANA, TEL European Library or the portal of the Consortium of European Research Libraries (CERL-MSS), as well as well-known search services for research information resources like EBSCO Discovery Service and The Summon® Service. In practical terms it means that the data of any partner of ours, be it even a small regional museum, automatically become the constituents of the major research and cultural information centres of the world.

From this standpoint it would be interesting to see how these global services or portals find themselves among the most significant generators of the traffic of Manuscriptorium. If we analyze the website visit statistics we will see that the majority of users address us directly (1/4 of the traffic) or come from Google (this group is not so numerous but still constitutes another1/4 of the traffic). Among the others the major are EUROPEANA, websites of partner organizations, Czech search engine seznam.cz, Wikipedia and Facebook.

However, if we look at the traffic generation from another angle, specifically from the angle of reference pages, i.e. when a user comes directly to study the document of interest, then ranked first is EUROPEANA with a share of 27.5%, ranked second is our National Library, ranked third is Czech Wikipedia, and ranked sixth is Facebook; what is more, when taken together, reference pages generate almost half of the total traffic.

The presence of such traffic generators as Facebook or Wikipedia shows that users make active use of the content of Manuscriptorium, and that this content creates their intention to split the information obtained. For us, this is the evidence of the fact that it makes sense to improve the website customization tools continuously.

In the last year, Manuscriptorium has been visited by users from 174 countries; almost half of them have been from foreign countries. According to the number of website visits the countries are ranked as follows: Germany, Poland, the US, Italy, Austria, Spain, France, Slovakia, Great Britain, Romania, Russia, Ukraine, Canada and the Netherlands. They are followed by Mexico, Japan, Brazil, Argentina and other countries from all continents.

Manuscriptorium is a very special resource with a special kind of users – by no means they all can read and study old manuscripts and books. Manuscriptorium

was created as a virtual research environment and has been developing in this direction. Nevertheless, according to the statistical data on access to the information resources of the Czech Republic National Library, Manuscriptorium has more searches or sessions and more browsed documents or objects than all our licensed (purchased) electronic resources taken together. By number they exceed the volume of Manuscriptorium by several times and include documents in modern languages, i.e. they are available to a large number of users (for example, ebrary has more than 120,000 scholarly books in English).

No doubt the seamless aggregation philosophy has some drawbacks: technological (not everything works in real time without failures at our partners'), political and cultural ones (for example, not everybody wants to share his/her resources; not everybody wants to cooperate with us; some organizations block open wide access to digital copies of the treasures from their collections). The most important requirement for cooperation is reliability of the partner. It is essential that every partner either discusses with us or at least advises us of all amendments that change the rules of Romanization of the aggregated content and records changes in the location of the referenced data from their digital storages in the OAI profile which serves as a medium for aggregated resources.

Our library signs an agreement with each Manuscriptorium's partner. This agreement stipulates the owner of the data and metadata (they remain in the partner's ownership) and what we are authorized to do with them (generally speaking, this is all about collecting and indexing of metadata and remote manipulating of images for the demonstration of the content); in other words, we purchase a license for all this, and we have to bear in mind that these are not we but the users of our digital library who manipulate with the images. Images are added to our interface in response to user demands. These demands are reflected in the user's behavior and instructions he/she gives in the course of the session within Manuscriptorium.

All the content in Manuscriptorium is in open access. It goes without saying that we are trying to attract new partners. For instance, recently the National Library of Armenia has become our new partner. We incessantly strive to improve the operation of our digital library by way of conducting new application research and introducing new functions and services. Not all of them have been formally included in the official version of Manuscriptorium, many are still working in the pilot trial mode.

Certainly, on a regular basis, we add content which is a result of digitization of our library collections. In the near future, in addition to new digital copies of our manuscripts, we can expect mass introduction of digital copies of dozens of thousands of books of the 17th and 18th centuries from our joint project with Google.

According to the agreement with our library, Manuscriptorium's administrator is "AiP Beroun", a Czech company. However, there are two organizations which are responsible for the operation of Manuscriptorium, while the funds for its development and faultless work are provided by our library with the help of the Ministry of Culture. At present, in addition to the official Manuscriptorium version (No. 2 – http://www.manuscriptorium.eu), the new Manuscriptorium is working (No. 3 – http://v3.manuscriptorium. com). This new version has been devised with due account of criticism which was expressed by our users. It works with the same content as the official version, but contains a lot of improvements, in the first place, in the field of using the content and its customization. After testing the new version all additional modules (or their replacements), to which our users have got used to, are put into operation, and the new version appears on the website of the official version. Good news is that this new version can work with all scripts not in the display mode only but in the search mode as well. The complete interface of both versions is in Czech and English.

We will be happy to start collaboration if you are interested in becoming a partner in Manuscriptorium.

Anatoly ZHOZHIKOV

Director, New Information Technologies Centre,
North-Eastern Federal University
(Yakutsk, Russian Federation)

Svetlana ZHOZHIKOVA

Leading Programmer,
New Information Technologies Centre,
North-Eastern Federal University
(Yakutsk, Russian Federation)

Indigenous Minorities of the North in Cyberspace: Experience and Prospects

In this paper the authors disclose (a) the ways of preserving and developing the languages and cultures of the indigenous minorities of the North in cyberspace by means of digital media and (b) the respective challenges of these processes. Introduction of specific fonts of the indigenous minority languages and creation of an Internet portal dedicated to these languages are examined.

On the verge of the new millennium we stepped into the century of information society, i.e. a period of establishing an aggregated sociocultural society as a basis of dialogue and interaction of civilizations, cultures and religions, a period which depends largely on the actual usage of information and communication technologies.

The rapid development of the Internet became an impetus for the development of social networks. As a result, there appeared a new multicultural communication environment devoid of borders, distances and time limits. However, it brought along a very serious problem, i.e. a possibility of obtaining Internet information and services in the dominating languages only. For instance, languages that are not presented on the Internet (one may find there only 400 languages out of the 6,700 existing in the world) cannot participate adequately in the information exchange and have to live in the shade of the "dominating nations" which, by imposing their languages, impose also their views of life and customs. These factors accelerate significantly the extinction rate of the minority languages and cultures.

At the same time, provided the opportunities are used reasonably and correctly, the Global Web is capable of becoming an integrator in the process of building a new world order and opening up new opportunities for not only preserving the languages of the peoples living on our planet but also contributing to their development.

The Russian Federation is a multiethnic country like many other countries of the world. Its territory is home for over 160 ethnic groups, 45 of which are indigenous minorities. 40 out of these 45 pertain to the indigenous minorities of the Russian North, Siberia and the Far East.

Every language is a unique information storage of the respective ethnic group. It reflects this group's culture, evolution and nature of humans as biological species. Throughout the centuries-old history of humanity some languages have appeared, but some died. However, the extinction process has become really critical recently. According to the pessimistic forecasts, by the end of the 21st century there will remain only about 10% of the currently existing languages. Therefore, urgent measures have to be taken to preserve linguistic and cultural diversity in both the Russian Federation and the world over.

Indigenous minorities of the North are the creators and keepers of a unique human culture. They represent a significant part of modern civilization. They have always been the carriers of adaptive life support systems in the severe conditions of the Arctic Regions and Extreme North, unique original traditions and specific spiritual values. For centuries, residents of the North have been mastering Arctic landscapes, adapting to extreme environmental conditions in permafrost, developing their original culture and living in harmony with nature, striving not to harm the vulnerable northern ecology, but to preserve it. However, indigenous minorities of the North are facing standard challenges of modern society more than other peoples of the world: the deepening globalization and technological development together with the active industrial exploration of their habitat have seriously impacted their traditional way of living. Rapid globalization processes and industrial exploration of the North have put these peoples on the verge of extinction. As was noted by the 4th Indigenous Minorities Congress, this threat is real for 12 out of 40 indigenous minorities of the Russian North, Siberia and the Far East.

The Universal Declaration on Cultural Diversity adopted by UNESCO on 2 November 2001 states that for the diversity of cultures, tolerance, dialogue and cooperation, in a climate of mutual trust and understanding are among the best guarantees of international peace and security. As a source of exchange, innovation and creativity, cultural diversity is as necessary for humankind as biodiversity is for nature. To preserve and develop this unique heritage it is vital to preserve in the digital form the language and culture of every people, represent them on the Internet and thus ensure their presence in the global information environment.

The analysis of Internet-based materials about indigenous minorities of the North revealed that there are no systemic Internet resources on this subject, and that the resources available are incomplete and insufficiently informative. A quite detailed analysis of these resources was made by A. Burykin in the article *Internet-resursy po teme "Yazyki malochislennykh narodov Krainego Severa, Sibiri*

i Dalnego Vostoka Rossii". Obzor imeyushchegosya materiala i pol'zovatel'skie zaprosy (Minority Languages of the Russian Extreme North, Siberia and the Far East. A Survey of the Materials Available and User Queries) [Burykin 2008]. Later on, the state-of-the-art of multilingualism in cyberspace in the Russian regions and Russia in general has been described in the collected works published by the Russian Committee of the UNESCO Information for All Programme and the Interregional Library Cooperation Centre. However, the situation hasn't changed drastically since then. Therefore, it has occurred to us to create a uniform portal about indigenous minorities of the North who inhabit the northeastern part of the Russian Federation (www.arctic-megapedia.ru).

New Information Technologies Centre of the North-Eastern Federal University started working on this task in cooperation with the leading scientists of the Institute for Humanities Research and Indigenous Studies of the North of the Academy of Sciences of the Russian Federation within the Development Programme of the Ammosov North-Eastern Federal University, project 4.1 «Preservation and Development of the Languages and Cultures of the Peoples of the Russian North-East», event 2.50, under the title «Preservation and Development of the Languages and Cultures of Indigenous Minorities that Are Presented on Digital Media and in Cyberspace».

Expeditions have been launched to the places of residence of these peoples in order to take videos and photos of the already vanishing language speakers and culture bearers (Yukagirs, Dolgans, Evens, Evenkis and Chukchi). The material collected is being used in two ways:

1. Production of digital educational multimedia DVDs on the languages and cultures of the indigenous minorities of the Russian North.

2. Creation of an open portal of the indigenous minorities of the North www.arctic-megapedia.ru.

The materials collected provided the basis for 17 content-rich educational DVDs dedicated to the languages and cultures of the indigenous peoples living on the territory of the Republic of Sakha (Yakutia). The www.arctic-megapedia.ru portal was devised to house virtually all the available materials. The thus issued DVDs were presented at several regional competitions and won the following awards:

- The Yukagir language and folklore textbook got The First Place Diploma in the nomination «Electronic Publication» at the 7th Interregional Exhibition and Fair «Yakutia's Printing House–2011» (June 2011).

- A set of the Yukagir language and folklore textbooks (5 discs) was awarded a silver medal in the competition «The Best Electronic

Publication» at the 15th Far-East Exhibition and Fair «Yakutia's Printing House–2011» (September 2011).

- The Arctic Multilanguage Portal www.arctic-megapedia.ru has become a finalist in the competition for the Far-East Internet Award «White Crane» in the section «The Best Nonprofit or Subject-Oriented Project of the Far East» (April 2013).

The results obtained show that this project is in high demand in the Republic of Sakha (Yakutia) and other regions of the Russian Federation where the indigenous minorities of the North reside (we have got cooperation offers from the Yamalo-Nenets Autonomous Okrug, Chukotka and Khabarovsk Krai). In March 2013, the project was presented at the 7th Congress of the Indigenous Minorities of the Russian North, Siberia and the Far East, held in Salekhard, and attracted great interest and support.

Generalizing the aforesaid, we believe that it is necessary (1) to accelerate the work in this field since the number of the indigenous language speakers and culture bearers is reducing, (2) to preserve this information in the digital format and (3) to present it on the Internet. This will allow us to not only preserve but also infuse blood into the languages and cultures of the indigenous minorities of the North in the global information environment.

References

1. Burykin A. (2008). *Internet-resursy po teme "Yazyki malochislennykh narodov Krainego Severa, Sibiri i Dalnego Vostoka Rossii". Obzor imeyushchegosya materiala i pol'zovatel'skie zaprosy.* In: Kuzmin, E., Plys, E. (eds.) Yazykovoye raznoobrazie v kiberprostranstve: rossiiskii i zarubezhnyi opyt, pp. 111–129. Moscow, Interregional Library Cooperation Centre, 2008. (in Russian)

2. Mnogoyazichie v Rossii: regional'nie aspecty. Moscow, Interregional Library Cooperation Centre, 2008. (in Russian)

3. Kuzmin, E., Parshakova, A. (2011). Razvitiye mnogoyazychiya v kiberprostranstve: posobie dlya bibliotek. Moscow, Interregional Library Cooperation Centre, 2011. (in Russian)

4. Zhozhikov, A. Zhozhikova, S. (2013). *Yazykovoe i kul'turnoe raznoobrazie korennykh malochislennykh narodov Severa v vek tsifrovykh tekhnologii.* In: Nikiforova, L., Nikiforova, N. (eds.). Nauki o kul'ture v perspektive «digital humanities»: Materialy Mezhdunarodnoi konferentsii, 3–5 oktyabrya 2013, Sankt-Peterburg, pp. 438–441. Saint Petersburg, Asterion. (in Russian)

SECTION 2.
SOCIO-CULTURAL ASPECTS OF LINGUISTIC DIVERSITY IN CYBERSPACE

Katsuko TANAKA

Assistant Professor, Nagaoka University of Technology (Nagaoka, Japan)

Understanding Social Phenomena in Cyberspace: Focusing on Language, Infrastructure and Contents[88]

1. Introduction

In this communication, we would like to show an example that is useful for understanding social phenomena in cyberspace focusing on three important components: human/substratum factor and products, media which connect all the components, and social systems/environment. The objective of this communication is as follows: whenever we see cyberspace phenomena, we tend to limit our focus narrowly to the substratum factor. Of course, we cannot be the Internet user without the substratum factor. But what matters for cyberspace phenomena is not only of the substratum factor but also the second factor of products themselves, and the third factor of consuming/producing activities of the products. And the centre of these activities is *human*. So, we need to understand all cyberspace phenomena from the *human-centred* perspective.

In cyberspace, main human activities are related to *literacy*. This is due to the following reason: most communications in cyberspace are verbal because human beings are not good at transferring all of non-verbal information verbally, or literally. And therefore, it is very important to understand the role of verbal language in cyberspace as an integrated system consisting of the three factors.

To construct an integrated system, we propose a new concept of *e-Network* introduced by Nakahira [2012].

2. Three Components for e-Network

The first step for constructing an integrated system is to define components, or factors. We set three factors for *e-Network.*

Human factor. The human factor includes the components attributed to human activities, and in the context of e-network it refers to human intelligent activities, represented by human users. This is the most important component in the e-network framework that creates dynamics of producing/ consuming information contents or social system, invoking the interactions between the human–human or human–substratum factors.

Substratum factor. The substratum factor plays the role of a device for the human factor to perform some action. It interacts directly with the human factor. It consists of the clients that produce/consume products, servers that stock/consume products, and the Internet that circulates products. It is introduced and operated in a manner depending on the environment such as the social system or customs developed by human beings.
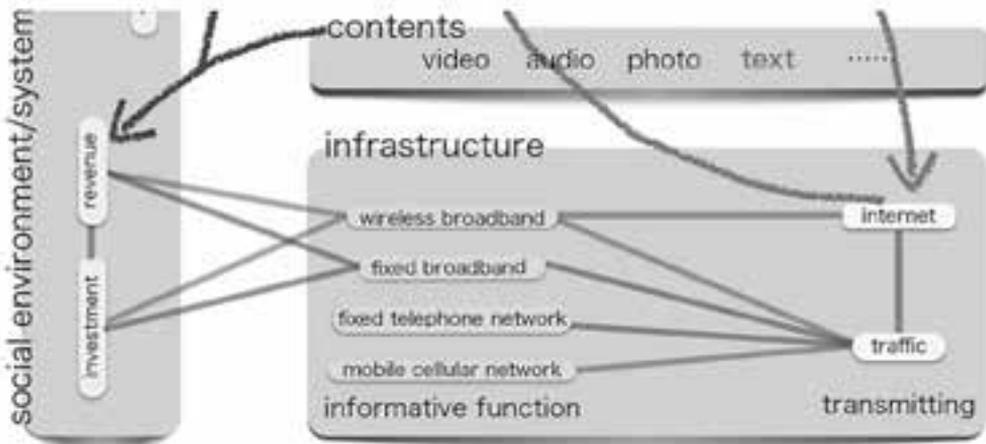

Figure 1. *Microstructure for the substratum factor*

Products. It is the information contents produced as a result of interaction between the human factor and the substratum factor. It consists of web pages, emails, software, and so on. Consumption of products might trigger interactions between the products and the human factor and between the products and the substratum factor. It can also contribute to the evolution/innovation of the substratum factor.

These components are connected by *media*. It is the device which connects all the three of them. It interacts semantically with the human factor, symbolically in connecting the human and the substratum factors, and in an encoded form with the substratum factor.

With these components, we can understand many cyberspace phenomena easily and clearly. For each component, detailed descriptions will be provided based on observations and thought experiments.

3. Substratum Factor

The easiest component to observe is the *substratum factor*. Many institutions or organizations make statistical data for the hardware/software infrastructure. In these days, not only statistical data for hardware but also traffic (software) data are widely available. And therefore we can easily consider and construct micro processes for the substratum factor.

An example is shown in Figure 1. We use the ITU statistical data to construct the structure. The ITU data are classified into several categories. Using it, we analyze the relationships between these categories. The blue lines are for the 2002–2006 dataset, the red lines are for the 2007–2008 dataset. The categories can be divided in the human economic activities such as revenue to informative function and investment from informative function. All transmitting activities are carried as informative function. The Internet provides the function that mutually connects the components.
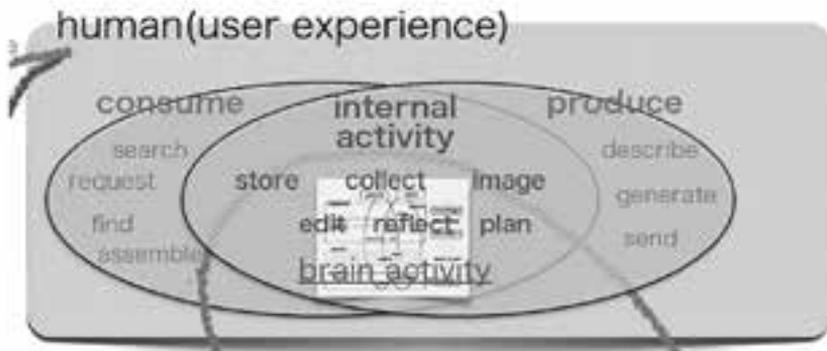

Figure 2. *Microstructure for the human factor*

4. Products

Next, we consider products. In cyberspace, we replace contents. In cyberspace, we transform contents. Products may be of two types: visual

and auditory materials. Visual materials consist of texts, photos, pictures, figures, and so on. Auditory materials consist of voices, songs, and so on. Complex materials are created by combining visual and auditory materials, consisting of video, book with sound, and so on. We easily recognize rich materials that include rich (signal) *information*. It means that we need to prepare substratum that has sufficient characteristics to make it possible to distribute such rich materials. From these considerations, we can easily understand text-products are the easiest to distribute in cyberspace because it has necessary features, i.e., transmitting rich information with the least amount of signals, which are the primary nature of *language*. In this way, we derive an important issue: whether people can treat their mother language or not on a computer in the e-network would have a significant impact on how information is produced and consumed by them in the e-network.

5. Human Factor

The most complex is the *human* factor, because it has many roles. We consider the factor by taking the following steps. First, human beings in cyberspace are regarded as experiencing users. In the experience, we separate two activities: consuming and producing. Each activity is derived from the human internal activity as brain activity. We recognize these activities are carried out in our brain. The brain activity is closely related to *knowledge* transfer activity. So if we construct the internal activity process, we have to construct a knowledge framework.

Figure 3 shows the relation of contents, information, and knowledge. In short, if we get knowledge from contents, literacy has one of important roles to play. When we recognize materials as contents, we start to translate them to information by using *literacy for pull*. The translation will be done internally in the brain, and we cannot observe it. If afterwards a person starts to collect and store information, it is just mere "information". But if one starts to organize the information they stored or collected, it will be translated into knowledge. Then, one m a y just keep the knowledge in their brain, but may also decide to provide their knowledge to someone who wants to get it. In this phase, a human uses literacy for push, including language. In the process, the language plays an important role in the view point of producing contents. Without language, a human cannot make any activities of consuming/providing products on the *e-Network*.
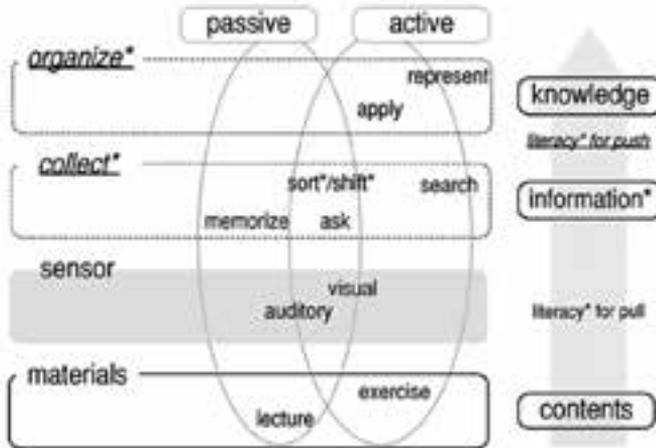
**Figure 3.** *Microprocess model for knowledge from Nakahira et al. (2014)*
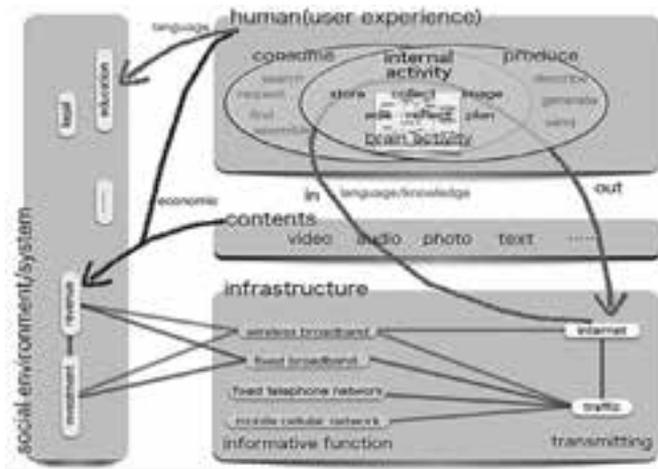
Figure 4. *Microprocess model for e-Network*

6. Connecting These Components with Media

We consider three components in the context of *language*. Any components are related with language, so we connect these components using language. In the Internet, most packets include signals for language, and when a computer receives the packets, it tries to convert them from electric signals to language materials, namely *products*. When a human detects the products as information, their brain will do several activities for collecting or organizing the information. These activities are regarded as those of consuming/producing of contents.

And these activities also generate social systems or environment, such as economy, education, and so on. Through revenue or investment next distribution cycle of products will be generated. Figure 4 shows the framework.

7. An Example for the Cyberspace Phenomena: Viral Media

Now, we are ready to understand cyberspace phenomena. As an example, we would like to apply the framework to understand a phenomenon in cyberspace – viral media.

Let's make a thought experiment. Assume that some Internet users have enough literacy for push by using the Internet function to communicate each other. Namely, they have sufficient education and rich legal/economic environment. However, the substratum they can use at that moment is not sufficient for treating rich products. One day, they come across a function which makes it possible for them to treat and share rich media with great ease. They start to use the function and reprocess the products with languages or computer techniques they have. The reprocessed products are light enough for some poor substratum to process easily; including not only poor infrastructure but also mobile environment. As a result, many people who have mobile environment may get the reprocessed information, and some persons who have interest in the products may start to use the new function; a kind of avalanche phenomenon. Other users may be eager to get a new piece of mobile hardware to consume the reprocessed information and produce another. Once the process starts, reproduction of information will spread virally. And all events develop based on language.

In this sense, there is a possibility of divide, not only digital but also knowledge one. In such a situation it is important whether one can use his/her mother language or not. For this reason, we need to continue monitoring cyberspace phenomena using "language".

References

1. Nakahira, K. T. (2012). A framework for understanding humane-network – interactions among language, governance, and more. Presented at the III Symposium international sur le multilinguisme dans le cyberespace. http://www.maayajo.org/IMG/SIMC/paris-v2.pdf.
2. Nakahira, K. T., Watanabe, M., and Kitajima, M. (2014). Assessment of developmental stages of generic skills: A case study. In: *Proceedings of the 22nd International Conference on Computers in Education*, pp. 200–205.

Galit WELLNER
Research Fellow, Ben-Gurion University
(Tel-Aviv, Israel)

The Importance of Multiculturalism
for the Flourishing of Human Beings

Why is it important to promote multilingualism and multiculturalism? In this paper I will try to provide one possible answer. In a nut shell, I will argue that multiculturalism can significantly contribute to the well being of the people who practice such multiplicity. Multiculturalism produces this effect by cultivating the sense of belonging. Belonging to a certain culture and society is a significant force in our everydayness. The feeling of cultural belonging is an important ingredient in our well being.

A preliminary question in this context would be what is culture to which one may feel a sense of belonging. Culture can be understood in two ways.

The first is the traditional understanding in which culture is a given that operates on the members of the society through established institutions. People in such a culture are expected to accept its rules and norms, obey the institutions and operate accordingly. Culture is dictated by institutions. These institutions can be the theater, the ballet, the library, etc. It is easy to detect them. Whatever is done outside them is not considered culture, let alone "good culture". In the worse version of this type of culturalism, these institutions are not open to other cultures; they tend to one cultural form and are likely to protect it.

The second model conceives culture as an endless process, dynamic and curious. In this process members of the society and culture affect each other as well as the culture in which they operate. It is a process of co-constitution of humans and their cultures. People do not need to obey, to be subjected to cultural institutions. Instead, they regard those institutions as equals, not as superiors. Because of the complexity of the co-constitution, in this context culture comes in the plural and is widely construed.

This second model of *Culture* has been developed by anthropologists since the end of the 19th century and has significantly evolved during the 20th century. According to this model *Culture* is a way of living; it is the social legacy that individuals acquire from their social group; it is a way of thinking and believing; it is a way of acting and performing; it is a socialization system, and a warehouse

of learning options; it is our past and our future, a system of symbols and their significance by which the social interaction is conducted.

Such a wide conception of culture relies on the basic view of Western Humanism that aims at the development of the *individual's unique personality*. This basic view allows the individual to have meaningful life, evolving through an on-going open-ended *exploration in* – and *a dialogue with* – cultural components and others' *Cultures*.

The foundations for this view were laid at the end of the 18th century and the beginning of the 19th by a group of British and (mainly) German scholars, writers and poets, including (on the British side) Shaftesbury, Byron, Mill, and (on the German side) Goethe, Schiller, Schlegel, Novalis, Holderlin, Humboldt, to name a few. Two main concepts were developed in this period: "*Bildung*" and "Kultur", or culture.

The concept of Bildung refers to "the process and 'product' of *Self fulfillment*" in which the individual is *proactively* developing and creating his or her personality through an explorative voyage through *Culture*. While originally it was mainly aimed at an adolescent young man (i.e. *Émile, or On Education* by Jean-Jacques Rousseau), today it is based on permanent transitions of everybody at any age. *Culture* – as stemming from its etymology – was conceived as a cultivating ground on which "life plans" and "life styles" can be flourished. Life is like a Petri dish in which culture is the nourishing basis, and Bildung is the process that develops in it.

Bildung, according to this view, leads to a sustainable experience of meaningfulness in life which *is an important condition of well being*. Gaining well being is obtained through a stable feeling of belonging to – and the acceptance by – something "bigger" than the self: community, society or culture. This process of achieving well being is expected to develop properly in a culture that is open to other cultures and is open-ended in the sense of avoiding determinism or teleological discourses.

What Was the Motivation for the Development of This View of Well Being?

The need for this view emerged at the second half of the 18th century. As secularism started to spread among intellectuals, they were looking for a view leading to a good and happy life as an alternative to the religious one.

They were well aware that for this to be possible, *Culture* had to be conceived as open and open-ended, that belonging to culture should not require a high level of rigid conformity, and that the relevant society had to allow the optimum of freedom, security and economic welfare for an individual to be able to explore

his or her culture along other cultures as well. They realized that culture should enable as many points of access as possible to different kinds of experiments and ways of living. Such a culture is multiculturalism.

How Can Culture Contribute to Well Being?

I propose to view the link between multiculturalism and well being through the prism of Self Determination Theory (SDT), a major theory in contemporary psychology that was developed by Richard Ryan and Edward Deci. SDT is the study of the conditions that foster (or undermine) positive human potentials. In exploring these conditions, they stress the major role of society in the flourishing of human beings ("SDT is concerned not only with the specific nature of positive developmental tendencies, but it also examines social environments that are antagonistic toward these tendencies").

Ryan and Deci portray a certain figure of well being that is beyond short-term happiness and hedonistic enjoyment. They write: "The fullest representations of humanity show people to be curious, vital, and self-motivated. At their best, they are agentic and inspired, striving to learn; extend themselves; master new skills; and apply their talents responsibly." They have identified some features that are general and applicable to all humans, no matter where they are, how old they are, and to which culture they belong. The theory of SDT assumes that all people share three groups of basic psychological needs: **autonomy** (in the sense of being able to direct oneself), **competence** (in the sense of being skilled) and **relatedness** (in the sense of belonging and acceptance). It can be phrased also as "I want to do", "I can do", and "I am accepted".

The first need for autonomy encompasses the need to feel that the main actions in one's life reflect and correspond to one's basic needs, tendencies and values. It is the need for an authentic self expression, for independence, for meaning and for choice.

The second category is the sense of competence. It comprises the need to experience oneself as being able to execute, to achieve, to turn one's dreams into a reality.

And the third category consists of sense of belonging and acceptance. This category incorporates the need be surrounded by loving and caring people, the feeling of being accepted "as is". The complementary set of feeling is that of being part of a group, of a society, of a culture. These two sets of feelings are

complementary to each other. Sometimes they are termed as empathy. Not only the receiver of empathy flourishes, but also the giver![89]

Culture and Language

So far I explained how culture operates as an essential factor in the forming of identity. Now, within culture, language plays an important role in building one's identity. The importance attributed to language has culminated in the 20th century, in what is known in philosophy as the linguistic turn. Wittgenstein, Heidegger, Foucault, Barth and many others, all have discussed the constitutive function of language in human life.

Daniel Everett writes: "All human abilities, including language, derive from two sources – genes and environment. The idea that language is exclusively a product of our culture or social environment is as simplistic, unhelpful, and wrong as the opposite idea that language grows like hair, shaped by our genome with no significant learning involved." [Everett 2010]. The tight link between language and biology should not be construed to lessen or minimize the importance and tightness of the link between language and culture. In this paper I will obviously focus on the link to culture.

We are born into a language that reflects a culture in the historical sense. We accept it as ours although we did not form it nor shape it. Yet it is "ours" in the deepest sense. The importance of language is not only "backwards" towards the past but also forward looking towards the future. When we start using a language, it shapes the way we view the world. Let us compare a language in which there are 10 words to describe various forms of lakes, of various sizes, to a language that has 10 words to describe forms of sand, some which move in the wind, some which are more solid. These two languages reflect two different sets of experience of the surroundings, of weather conditions and geographies that have shaped the local societies for ages. Each of this set of words forms a warehouse of metaphors with which the local culture operates. Another example is the naming of colors. In French and Hebrew (two languages that I happen to know) the color of the sky in the summer has one clear word – azure in French and tkhelet in Hebrew. In English the closest meaning requires two words "light blue". Everyone who visited London knows that the color of the sky there tends to be grey and is only rarely light blue. Even the thesaurus reflects this tendency, with the color blue having 12 synonyms, and grey – 29!

---

[89] *This last category represents the social aspects of being happy and self motivated and stresses the importance of relationships to the society and culture in which one lives. Philosopher Daniel Haybron wrties: "'all you need is love', the Beatles told us, and they weren't too far off the mark." He adds: "To say that relationships matter for human happiness is like saying water matters for fish" (p. 68).*

More than double! All this has an effect on one's identity. In other words, one's sense of belonging is shaped by the language which in turn is shaped by the environing conditions.

Daniel Everett conceives languages as "tools that fit their cultural niche" [Everett 2010]. He admits he is not the first to claim that language is a tool. It was already stated by Aristotle, and in modern times by Lev Vygotsky. He even roots this conceptualization in the biblical story of the Babel Tower, in which language was a tool used by the constructors of the tower to communicate with many people of various origins. Everett's concept of language as a tool is limited to communication purposes.

My few examples prove this claim, but I want to extend it and develop it into a co-construction of language, cultures and human beings. In so asserting, I am inspired by Martin Heidegger's famous statement in his book *On the way to Language*. Heidegger says that Language "is the House of Being". We dwell in our language. Language is the mental home to which we belong.

Such a broad conception of language allows Heidegger to further maintain that language shapes our thinking (Heidegger, *Identity and Difference*, p. 38: "thinking receives the tools for ... self-suspended structure from language"). This approach is more radical because it refers to language as a tool for thinking, when one communicates with oneself. Why is it more radical? Let's examine the writings of another philosopher, a French one.

Roland Barthes makes in the book *The Pleasure of Text* an interesting distinction between *texte lisible* and *texte scriptable*. The first is "readable" text, a text that brings pleasure, which is "fun". The latter is "writable" text, which challenges the reader's position as a subject. Barthes argues that "writable" texts are more important than "readable" ones because in them the text's unity is forever being re-established by its composition and by the codes that form and constantly slide around within the text. The reader of a readable text is mostly passive, whereas the person who engages with a writable text has to make an active effort, and even to re-enact the actions of the writer himself. In the terms of German romanticism, writable text is the part of language that participates in the *Bildung* process. In Heideggerian terms, writable text is a tool for thinking.

How Can Digital Reality Support Cultural Exploration?

Some scholars argue that today, in the age of the Internet, the Bildung process became virtual. For example, Sherry Turkle in her book "Life on Screen: Identity in the Age of the Internet" (1997) claims that today identities come in

plural. People are "cycling through different identities" (p. 179), as a matter of mainstream and not of margins. *Bildung* is not so much about travelling to various countries or reading many books. It is more about wandering in virtual spaces. She writes, "the Internet has become a significant social laboratory for experimenting with the constructions and reconstructions of self that characterizes postmodern life. In its virtual reality, we self-fashion and self-create" (p. 180). Turkle explains that we construct our identity in the interaction with the computer screen (cell phone included), instead of a face-to-face interaction with other people. However, one should realize that some of the activities in front of the screen are with other people, via emails, chats, posts, etc., what is known as computer-mediated communications. In such interactions, people tend to play with "sets of roles that can be mixed and matched, whose diverse demands need to be negotiated" (p. 180). Computer mediated communication contributes to the reconstruction of identity as multiple.

The digital sphere contains an infinite amount of information and it is a medium which supports and sustains interaction between masses. These two components for themselves can support and encourage cultural exploration. It seems that the new digital space we live in almost automatically guarantees all conditions necessary for exploration in the above sense: infinite ability to experiment in various ways, of many cultures.

In an environment that is based on multiplicity, it is natural to support more than one language. Multilingualism can be easily practiced on the Internet, both technically and psychologically. It is our mission to ensure that it is also feasible socially and culturally. We have to ensure that the underlying cultural diversity is also maintained.

How can we ensure that our digital technologies enable cultural diversity? Here I would like to implement Bernard Stiegler's concept of epiphylogenetics. Stiegler shows how technologies have always developed in parallel and in close relation to human evolution. The human and the technological, he posits, are entangled, and one influences the other in an endless process. It is a process of co-constitution and co-shaping. He shows how in pre-history, the usage of simple tools like a hand-axe, contributed to the development of our brain's cortex and to our upright bodily position. We had to free the hands and construct the necessary neural circuits. The human and the technological were created at the same moment. Stiegler terms his concept "epiphylogenetics" as a combination of epigenetics and phylogenetics: if genetic is about the inheritance of genetic code, epigenetics is the genetic changes that happen outside the DNA, those that

lead to changes and differentiations. Phylogenetics is the study of the history of these changes. Technology, Stiegler argues, is the human memory that exceeds the genetic memory; it is the exteriorization of human memory.

The process of epiphylogenetics never stops. It continues each and every moment, as we use ancient and modern technologies alike. When we wear clothes, when we eat cooked food, when we work in front of a screen and when we use language. All these are technics and technologies that construct us as humans. This construction – or better co-construction – goes on endlessly. All these technologies participate in the construction of self-identity, in *Bildung*. They are not only ways to explain our past and present, they also shape our futures.

Today, the processes of co-construction are obviously much more complex. In a simplified way, maintaining a multi-cultural environment is likely to encourage people to use the digital tools they have in creative ways that will support multilingualism. The reverse is also true: the more our tools enable multilingualism – for example by easily switching from one language to another – the more languages will be used, maintained and kept updated. Moreover, our technologies should support multilingualism in a way that reflects the various underlying cultures, so that words with multiple meanings should be given more than one possible translation. Such linguistic richness should be preserved.

It is our responsibility as society to ensure that we enable multiculturalism. It is also our responsibility to encourage the developers of technologies to enable it as well. Perhaps more importantly, our technologies should enable a dialogue between languages and cultures (cf. Heidegger, *On the Way to Language*, p. 5). When we learn a new language, it is important to learn also its cultural background.

Summary

Linguistic diversity is one of the main pillars of cultural diversity and hence the importance of fighting language marginalization. In this paper I wish to ask not only what can be done to minimize language marginalization but also what is the rationale for such a strategy. It is a shift from a "how" question to a "why" question.

I believe that a goal should be set for the preservation of linguistic diversity. This goal can direct the efforts of preservation and possibly make them more effective. The proposed goal is well being, the well being of those who speak marginalized languages. Our claim is that language preservation is important

for the well being of those who speak it because language is one of the important markers of cultural identity. If one can identify herself or himself with their culture, including speaking their language, then she or he is more likely to flourish and sustain the challenges life poses for us, in developed and developing countries alike. Take for example an economic crisis, accompanied by high unemployment rates. Our hypothesis is that those members of the local society whose well being is better maintained will be able to cope with the crisis in a more effective way, their level of happiness will be higher, and they will find solutions that will help them restore their situation.

The preservation of marginalized languages can also contribute to those who speak other languages, as marginalized languages can offer new perspectives on the world, produce new metaphors and enrich our thinking as a multi-cultural collective.

Digital technologies are able of assisting us in this important mission. They can be adjusted to various languages in a relatively short time and with relatively limited resources. They can bring the voices of marginalized languages to each and every corner of the world. And their impact can be fast, as we saw from the introduction of various Internet-based technologies to our everydayness.

Vicent CLIMENT-FERRANDO

Associate Professor,
Department of Translation and Language Sciences,
University Pompeu Fabra
(Barcelona, Spain)

Diversity Advantage: Migrant Languages as Cities' Social Capital. Barcelona and London Compared

Abstract

Migration and human mobility have experienced an unprecedented increase over the past few years. For the first time ever in 2010, the majority of the world's population was predominantly urban, and this proportion continues to grow. By 2050, more than 70% of the world's population will live in a city. This high concentration of people in urban settings has resulted in an increasing linguistic diversity, a trend that will continue over the next years. The goal of this communication is to analyze if and how 21ˢᵗ century cities are considering linguistic diversity as a social and economic value for development. I will analyze London and Barcelona to provide a detailed description of how these two cities are approaching linguistic diversity derived from migration.

Introduction

International mobility and migration has experienced a rapid increase over the past few years. The latest 2013 figures of the United Nation's International Migration Report confirm this trend: between 1990 and 2013, the number of international migrants worldwide rose by over 77 million or by 50%. Much of this growth occurred between 2000 and 2010. During this period, some 4.6 million migrants were added annually, compared to an average of 2 million per annum during the period 1990–2000 and 3.6 million per annum during the period 2010–2013[90]. Put in other terms, approximately 1 billion of the world's

---

[90] *International Migration Report 2013 Available at: http://www.un.org/en/development/ desa/population/publications/pdf/migration/migrationreport2013/Full_Document_final.pdf#zoom=100.*

7 billion people are migrants. Some 214 million are international migrants. Another 740 million are internal migrants.[91]

Migration historically has been the process by which different ethnic, cultural, language, religious groups have come into contact and thus presented both migrants and host communities with many challenges. In the contemporary era of globalization, the potential for such mixing has reached unprecedented levels so that the challenges of coping with diversity will increase over the years to come. Castles and Miller [2003: 14] have identified two central global issues, which have arisen from the mass population movements of the current epoch: the regulation of international migration on the one hand, and its effects on increasing ethnic, cultural and linguistic diversity, on the other.

This increasing diversity has become more evident in urban areas. While it is true that one of the traditional units of analysis have usually been nation-states, cities have increasingly become the hub to analyze and measure the current population trends. In the words of Adrian Favell: "The most promising development would seem to be the growing focus, not on comparing nation-states, but on comparing cities and the migrants who populate them. Cities are the arena where the newest and sharpest developments are first observed, and where there is a degree of cross-national convergence on both policy problems and policy solutions, that belies many of the differences reflected in national ideological debates. It is at this level also that the new research agenda on transnationalism and the 'globalisation of place' makes some sense. Paris, London, Amsterdam, Brussels, Berlin and Milan, among others, have become multicultural cities in ways quite unexpected and unintended by national governmental policy makers, and each requires attention to its specificities within a national, regional and international context, in order to explain how" [2001: XIX].

Urban growth has been experiencing an exponential increase over the past decades and will continue to do so in the years to come: one hundred years ago, 2 out of every 10 people lived in an urban area. By 1990, less than 40% of the global population lived in a city, but as of 2010, more than half of all people live in an urban area. By 2030, 6 out of every 10 people will live in a city, and by 2050, this proportion will increase to 7 out of 10 people. Currently, around half of all urban dwellers live in cities with between 100,000 k–500,000 people, and fewer than 10% of urban dwellers live in megacities (defined by UN HABITAT as a city with a population of more than 10 million). In absolute terms, from 3.5 billion urban dwellers in 2010, figures will jump to 6.4 billion by 2050.

---

[91] *Migration and the United Nations Post-2015 Development Agenda. Available at: http://publications.iom. int/bookstore/free/Migration_and_the_UN_Post2015_Agenda.pdf.*

1. Managing Diversity in Urban Environments: The Great Challenge of 21st Century Societies

Immigration and human mobility have entailed profound demographic but also economic and social changes, highlighting one of the main challenges that host societies currently face: managing diversity. It is at the local level – the closest sphere to the citizen – where the demographic and social changes become more palpable and it is, therefore, at the level where the vast majority of actions aimed at managing diversity are concentrated. "More action at local level" was one of the main recommendations by the European Agenda for the Integration of Third-Country Nationals[92] – establishing the European Commission's guidelines on immigrant integration – which states that local authorities play an important role in shaping the interaction between migrants and the receiving society. Cities are considered to play a determinant role in managing this diversity.

One of the latest reports on the State of the World's Cities (2012–2013) strongly encourages that today's people-centered cities of the 21st century stimulate local job creation, promote social diversity, maintain a sustainable environment and recognize the importance of public spaces. It is a city that is all encompassing and accessible to everyone[93].

Following the same discursive line, the Council of Europe's Intercultural Cities project[94] (2008–2013) has proposed a new way of thinking and action upon diversity analytical and conceptual tool aimed at exploring the advantages of the increasing diversity far from the negative, narrow stereotypes that have often been associated with immigration. While it is true that diversity – be it religions, cultural or linguistic – can generate conflict due to prejudices by the host society revolving around issues of overuse of welfare system or the job market, diversity can also create benefits as it increments the variety of goods, services and skills available in urban environments. The increased number of competences and skills provided by diversity can also foster creativity, innovation and economic growth [Berliant and Fujita 2003].

The different studies conducted recently confirm that a well planned strategy on managing diversity can lead to social and economic benefits for the society as a whole. The report *Evidence of the Economic and Social Advantages of*

---

[92] *Available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0455:FIN:EN:PDF.*

[93] *UN Habitat. Available at: http://www.unhabitat.org/pmss/listItemDetails.aspx?publicationID=3387.*

[94] *Available at: http://www.coe.int/t/dg4/cultureheritage/culture/Cities/ICCOutcomes_en.pdf.*

Intercultural Cities Approach[95] published by the Council of Europe provides an empirical analysis on how cities are best positioned to develop the capacities to steer the effects of immigration on the society by providing for increasing benefits of heterogeneous communities and reducing their negative effects and claims that while it is true that immigration will continue to be a burning topic, there is a necessity to develop proper public policy tools to provide for larger scale benefits of heterogeneous society.

2. Linguistic Diversity in Cities. Tower of Babel or Source of Opportunities?

One of the most visible changes in urban environments has been the increasing coexistence of different languages. Today, linguistic diversity in cities is the norm and not the exception. The 2011 population census[96] of the city of Toronto concluded that 48% of the people living in the city have as mother tongue a language other than English or French – Canada's official languages; very close in percentage to the city of New York[97], with some 47% of the population speaking a language other than English. Helsinki, with a much lower immigration rate – 7,2% l – hosts some 150 languages [Kraus 2011: 30] a bit less than the city of Manchester which hosts more than 200 languages, as highlighted by a University of Manchester research study[98], or London, with more than 300 languages spoken[99]. And the list of cities could be endless.

Unlike the traditional nation-based approaches and ideologies – which still revolve around the 19th century state-building one-language-one-state approach, cities go beyond the identity-based, nation-state building and can offer a much more accurate account of the vibrancy and heterogeneity of its people. It is the local level that offers a realistic picture of the real language landscape.

3. From Act to Action. Are Cities Taking Advantage of This "Diversity Advantage"? A Closer Look at Barcelona and London

While there seems to be a wide consensus in academia and in international organizations on the benefits of diversity, it still needs to be analyzed whether these theoretical postulates are translated into concrete, tangible city initiatives

---

[95] *Available at: http://www.coe.int/t/dg4/cultureheritage/culture/cities/research/literature%20review.pdf.*

[96] *2011. Toronto City Council: http://www.toronto.ca/demographics/pdf/language_2011_backgrounder.pdf.*

[97] *Language and Education in New York. Migration Policy Institute. Available at: http://www.migrationinformation.org/datahub/state2.cfm?ID=NY.*

[98] *Available at: http://mlm.humanities.manchester.ac.uk/aboutus.html.*

[99] *http://www.independent.co.uk/news/london-multilingual-capital-of-the-world-1083812.html.*

or programmes geared towards not only recognizing and making linguistic diversity visible but actively promoting it. This section will provide a succinct bird's-eye view of the linguistic diversity in two cities, Barcelona and London, to further analyze them in the subsequent section.

3.1. The Languages Spoken in Barcelona. A Bird's-Eye View

Over the past decade, immigrant population of Barcelona has multiplied by six, going from 3% in 2000 to more than 17% in 2013, a rapid and unprecedented change in a very short period of time, as shown in Table 1 below.

Table 1. Evolution of foreign population in Barcelona (2000–2013)
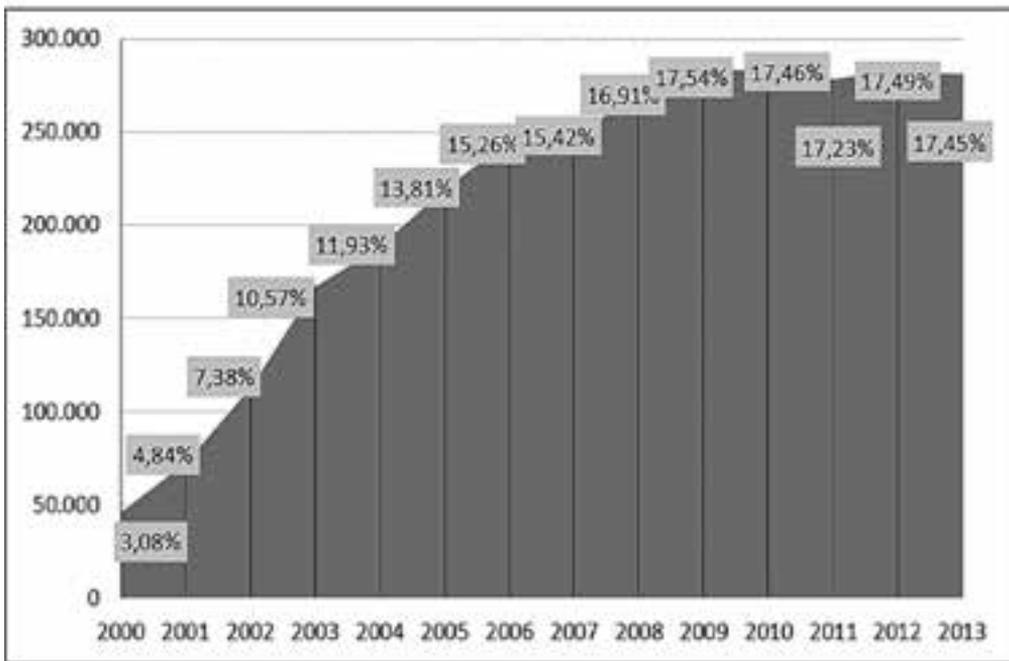
Source: Author's own elaboration from the Catalan Institute for Statistics (March 2014)

The new population living in the city comes from all five continents of the world, as shown in Table 2 below, which features the number of nationalities present in each neighborhood, the highest number being in the district of Eixample (145 nationalities) and the lowest Sant Andreu (115 nationalities), and Table 3, which indicates the main nationalities present in the city of Barcelona.

Table 2. Number of nationalities present in the city of Barcelona (2013)

District			Year			Nr. of nationalities

Districtos/Barrios	2009	2010	2011	2012	2013	Número de nacionalidades que coinciden
Ciutat Vella	14,3	14,5	14,7	15,2	15,7	139
Eixample	16,4	16,5	16,4	16,9	17,5	145
Sants-Monjuïc	12,1	12,3	12,5	12,9	12,9	131
Les Corts	3,3	3,3	3,2	3,3	3,3	120
Sarrià-Sant Gervasi	5,8	5,7	5,5	5,6	5,8	135
Gràcia	6,6	6,7	6,4	6,6	6,6	134
Horta	7,6	7,7	7,6	7,5	7,6	127
Nou Barris	9,6	9,5	9,7	9,8	9,5	123
Sant Andeu	6,5	6,7	6,8	6,7	6,7	118
Sant Martí	11,9	12,3	12,5	12,8	13,1	138
No consta	6,0	4,9	4,6	2,7	1,3	94

Source: Catalan Institute for Statistics (2013)

Table 3. Main nationalities present in Barcelona (2013)

Main countries of origin of immigrant population	
Country	**% of total immigrant population in Barcelona**
Italy	8,63
Pakistan	7,69
China	5,79
Ecuador	4,93
Bolivia	4,8
Morocco	4,73
France	4,52
Peru	4,5
Colombia	4,15
Philippines	3,11
Dominican Republic	2,7
Argentina	2,49
Romania	2,49
Germany	2,46
Brazil	2,24

Source: Author's own elaboration. Data from the Catalan Institute for Statistics (IDESCAT, March 2014)

When it comes to featuring the number of languages spoken by the new population, however, a less clear picture emerges as there are no official data available. We can only find the countries of origin but not their languages and we should avoid the one-state, one-language approach above-mentioned, which does not correspond to the linguistic diversity of states around the globe.

Therefore, we must use other sources of information to explore the linguistic diversity present in the city. To do so, I have resorted to two other complementary sources:

a. Mother tongues spoken by the immigrant student population. The Catalan Government's Department of Education identifies systematically the languages spoken at home by the students attending the so called *Aules d'Acollida* (Language immersion classrooms), aimed at immigrant students who have recently arrived and do not speak Catalan and/or Spanish. Aware that this only represents part of the linguistic diversity and not all, the Department has provided us with some information on the main 30 languages of these students (see Table 4 below), which can serve us as an approximate idea on the linguistic heterogeneity present in Barcelona.

Other than Spanish, we can see in the table that students have Arabic, Chinese, Tamazight, Punjabi and Urdu as the main languages spoken by immigrant students.

Table 4. Languages of immigrant students attending the "Language Immersion Classrooms" (2009–2010)

Llengües familiars dels alumnes que assisteixen a una aula d'acollida (Curs 2009-2010)			
1. Castellà	7284	16. Ucraïnès	133
2. Àrab	3089	17. Búlgar	115
3. Xinès mandarí/putonghuà	1600	18. Mandinga	108
4. Amazic	1193	19. Polonès	101
5. Romanès / moldau / valac	853	20. Ful	98
6. Urdú	673	21. Bengali	95
7. Panjabi	619	22. Hindi	74
8. Portuguès	482	23. Guarani	71
9. Soninké	261	24. Wu	60
10. Anglès	244	25. Twi	59
11. Tagal	225	26. Alemany	53
12. Xinès cantonès o yuè xiang	217	27. Bambara	53
13. Rus	214	28. Zhuang	50
14. Francès	208	29. Quítxua	48
15. Wòlof	180	30. Armeni	47

Pere Mayans, Departament d'Ensenyament i Mònica Fidalgo, Universitat de Barcelona

Source: Department of Education and University of Barcelona

b. The 2007 Demographic poll. We have also analyzed the poll carried out in 2007 in which it was asked about the languages, other that Spanish or Catalan, spoken at home. The results were revealing: more than 10% of those interviewed claimed that they spoke a different language, the most important ones being those contained in Table 5 below.

Table 5. Population with a language other than Spanish and Catalan

Language	%
Arabic	16,7
Romanian	14,3
Tamazight	12,2
French	6,5
Portuguese	6,2
Galician	5,1
English	4,7
Russian	4,1
German	3,8
Chinese	2,3

Source: Author's own source derived from the 2007 Demographic Poll

From the data gathered and analyzed, we can provide an in-depth account of the linguistic diversity present in the city of Barcelona: Punjabi and Urdu, the languages spoken by the Pakistani population; Mandarin, Wu and Cantonese, spoken by the Chinese population, Quechua, Guarani or Aymara, spoken by part of the Latin-American population, Arabic and especially Tamazight, spoken by around 50 to 80% of the Moroccan population living in the city; Tagalog, spoken by the Philippine population, Romanian, Portuguese, Russian, Italian, French and English. These would be, in very general terms, the main languages and communities present in the city.

3.2. The Languages Spoken in London. A Bird's-Eye View

London portrays diversity as one of its greatest strengths. The introduction to the Mayor's London Plan stated that London is 'the most culturally diverse city in the world'. It went on to make a strong case for international in-migration: 'London's diversity is one of its great historical social, economic and cultural strengths. New arrivals moving to London from overseas will contribute further to it. London is already a highly diverse city, one of the most multi-racial in the world. Nearly one-third of Londoners are from black and minority ethnic

communities... and a significant growth in (these) is projected over the next 15 years. International in- and outmigration has been high and is projected to remain so.' [MoL 2004][100].

In terms of the origins of immigrant population, Table 6 below shows the main nationalities present in the city:

Table 6. Origins of population in the UK

Western Europe	18%
Of which France	5%
Central/Eastern Europe	14%
of which Poland	5%
Australia/New Zealand	9%
North America	6%
of which US	5%
Caribbean	2%
Central/South America	5%
Of which Brazil	3%
Middle East	4%
South Asia	12%
Africa	19%
Of which South Africa	6%

Source: *The Impact of Recent Migration on the London Economy*, LSE (2007)

Unlike in the case of Barcelona, London – and the UK as a whole – has official figures on the number of languages spoken in the city. Statistics from the 2011 Census show that 78% of the capital's residents speak English as their main language (which does not mean they have English as their – only – mother tongue). But the remaining 22% – equivalent to just over 1.7 million people – have another first language. The most common other language is Polish, spoken as the main language by nearly 2%t of foreign residents in London, followed by Bengali, Gujarati, French, Urdu and Arabic. The most diverse borough is Hillingdon, where some 107 languages are spoken, followed by Newham, where 104 languages are spoken. Figures also reveal that some more 300 languages

---

[100] *See: The Impact of Recent Migration on the London economy. London School of Economics (2007).*

are spoken in 30 of the capital's 33 boroughs with only the City, Richmond and Havering falling slightly below this benchmark.

Using the 2011 Census, population geographer Guy Lansley illustrated the extent of London's linguistic diversity, identifying the main 80 languages, excluding English, spoken by the registered population, as shown in Table 7 below.

Table 7. Most commonly spoken language in London (excluding English)
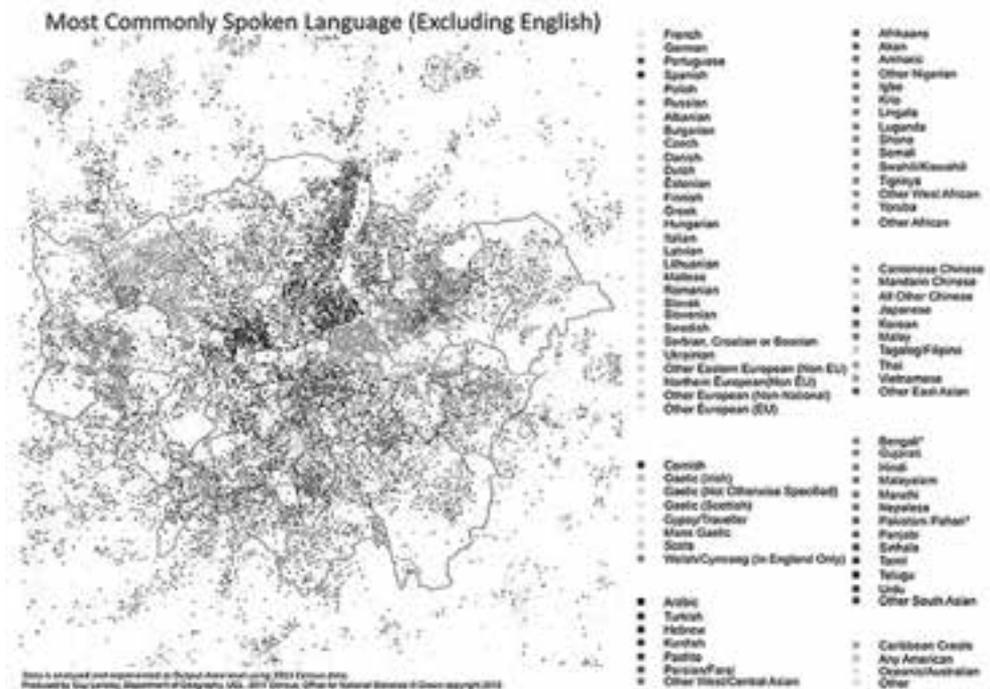
Source: 2011 Census language map of London

4. Urban Multilingualism. A City's Greatest Asset

Over the past few years, international organizations, researchers and national institutions have started to focus on the benefits of immigrant languages as a sign of modernity, openness and, above all, economic benefits derived from this increasing linguistic diversity. Back in 2007, the European Commission published the ELAN report, the Effects on the *European Economy of Shortages of Foreign Language Skills in Enterprise,* which highlighted the loss of commercial and economic activities by many European companies due to lack of language skills, many of them being the languages spoken by the migrant

population. Another interesting initiative is the campaign developed recently by the International Organization for Migration called *Migrants Contribute*[101] on the benefits and contributions migrant make to host societies, including languages; or the Council of Europe Conference to be held in Bilbao on 18–19 September 2014 on the importance of capitalizing the advantage of migrant languages in urban environments.

Back in 2005, the "Productive Diversity" approach followed in Australia [Pyke 2005] proved the tremendous benefits in capitalizing the languages and cultures of immigrants. Another recent term used in academia to refer to this fact is the "Diversity Advantage" [Landry & Wood 2008].

5. Are Local Authorities Aware of the City's Diversity Advantage? An Analysis of Barcelona and London

5.1. The Case of Barcelona: The 2012–2015 Immigration Plan

The current 2012–2015 Immigration Plan provides the guidelines on the immigration and social cohesion policies followed by the city, which place migrant languages as one of the core themes to be not only recognized but also actively supported and promoted.

Recognizing that "citizens' cultural and linguistic diversity is an added value that must be reinforced, as it benefits society as a whole", the Barcelona Immigration Plan has identified specific lines of action aimed at fostering the linguistic diversity of migrants. These are the following.

a. Mother-tongue instruction. The Plan wishes to work to increase the number of schools that teach immigrant languages through the "*Mother Tongue Instruction Programme*" of the Government of Catalonia's Ministry of Education, to consolidate the linguistic talent of students of different origins (action 50 in the Immigration Plan). The goal is to promote the latent talent for languages of youngsters who speak different tongues at home, as the potential plurilingualism involved would broaden their career opportunities. The plan states that "Bearing in mind the current context, marked by the financial crisis and the turn of the economic cycle, and in which we have to provide people with the resources they need to cope with growing competitiveness, we intend to create conditions in which public policies contribute to citizens' success in education and work, with particular emphasis on second-generation immigrants. We will promote training, talent, resource accessibility and

---

[101] *http://migrantscontribute.com/.*

a welfare society" (p. 42). In this respect, the authorities have ensured that despite the economic scenario, more mother-tongue instruction provision is available (see Table 8 below) so as to make sure immigrant languages are kept by the new citizens of Barcelona.

Table 8. Evolution of mother tongue instruction in Catalonia. 2004–2012

	Árabe	Amazige	Chino	Rumano	Holandés	Ucraniano	Bengalí	Portugués	Urdú
2004-2005	981								
2005-2006	981	81	80						
2006-2007	1121	81	71		144				
2007-2008	1200	91	129	180	130				
2008-2009	1596	109	185	465	125	35			
2009-2010	1596	134	217	308	15	58	25	227	
2010-2011	1682	98	390	204	137	53	57	227	20
2011-2012	2409	49	199	210	137	42	30	300	20

Source: Department of Education. Government of Catalonia

b. *The ProMES Project. Promoting Multilingualism in Exporting SMEs*
This is a European project being applying in different cities and regions across Europe aimed at exchanging best practices on the potential benefits of taking advantage of the languages of citizens in different cities. It is being carried out in Barcelona by the City's Chamber of Commerce and there are currently 10 auditors doing fieldwork and evaluating how companies are identifying and using the potential benefit of the city's multilingual capital. This project has been inspired by the above-mentioned ELAN project and concludes that the need for multilingual citizens will exponentially increase over the years to come. Despite not being a project directly mentioned in the Barcelona's Immigration Plan, the city and the Government of Catalonia have been actively involved.

c. Fostering of multilingualism in Barcelona's companies (Barcelona Growth project, action 54 of the Immigration Plan). The Plan refers to the imperative need to create a specific programme for encouraging

the use of the languages of emerging countries, so as to aid companies' internationalisation and improve the job prospects of those proficient in such tongues.

d. *Entrepreneurship and Migration (action 27 of Immigration Plan)* The Plan talks about encouragement for foreigners to participate in the programmes run by *Barcelona Activa*, a city agency aimed at fostering entrepreneurship among all citizens of Barcelona. The goal is to increase immigrants' involvement in training courses and provide with resources for entrepreneurship. According to figures, immigrants turn to be more entrepreneurial than local population in Barcelona; 23.4% of users are of immigrant origin while they represent 17% of the total population. Languages can play a key role when associated to the economy.

5.2. *The Case of London*

Diversity in London has often been linked in public rhetoric to the city's economic prosperity. In 2007, the City of London commissioned a report entitled *The Impact of Migration in London's recent economy*, which highlighted the positive effect of a diverse, heterogeneous population in terms of skills. The report mentions that "positive effects of migration are its *qualitative* impact on the London labor force and economy, through diversity, flexibility, international experience and skill sets; and its *quantitative* contribution through expanding labor supply."

Languages are considered as a positive asset in the report. The report states that there are different positive effects of having a diverse population, one of them being Facilitating trade relations with migrants' home countries, through exploitation of their language skills, market awareness, networks and social capital. The report, therefore, encourages the authorities to use this "diversity advantage" and put it at the service of the city.

Much more recently, in July 2011, the City of London published the overall strategic plan, the London Plan, setting out a fully integrated economic, environmental, transport and social framework for the development of the capital to 2031. Going through the Plan, we can also see several references to languages and the role they must play:

"London's diversity is one of its greatest strengths and one of the things its residents most appreciate about living here: more languages and cultures are represented in the capital than in any other major city. The Mayor is committed to securing a more inclusive London which recognizes shared values as well as the distinct needs of the capital's different groups and communities", and

continues highlighting the need for the city to "make the most of the benefits of the energy, dynamism and diversity that characterize the city and its people; embraces change while promoting its heritage, neighborhoods and identity; and values responsibility, compassion and citizenship."

While all official documents analyzed include linguistic diversity and migrant languages in their rhetoric as one of the city's greatest assets, we find no specific reference to tangible, concrete projects or initiatives developed by the local authorities that explain how these postulates are translated into concrete actions. It is in the sphere of academia and, especially, at grassroots level, however, that we can find a remarkable number of examples. As an example, a research project conducted by the London School of Economics (LUCIDE) has unveiled what is happening at grassroots level in education and migrant languages. As indicated in the project, London schools provide education for learners whose diversity is best expressed by the estimated number of languages they speak: 360. Inner London school population is over 50% bilingual (London Challenge Figures 2008). Some schools have over 90% bilingual pupils on roll. In addition there are 16 international schools and a growing number of bilingual schools such as: French Lycee, German Grammar School and Italian School, which provide bilingual and bicultural curriculum. In terms of complementary education, the estimates are that there are several hundreds of mother tongue schools providing support to bilingual children to maintain their home languages.

Concluding Remarks

This communication has highlighted the increasing interest in managing diversity at local level, and more specifically, in focusing on the languages of migrants as social and economic assets. The analysis of the two case studies, Barcelona and London, has pointed out at the following key findings:

1. Despite not having official data on the number of languages spoken in the city, the Barcelona Immigration Plan has placed migrant languages as one of the key elements in its immigration strategy, identifying concrete proposals and working with the Catalan regional authorities to capitalize the city's linguistic diversity.

2. Unlike the case of Barcelona, the 2011 UK Census contains detailed information on the languages spoken in the city of London, Polish being the most widely-spoken language in the city.

3. The 2012–2015 Barcelona Immigration Plan considers the languages of migrants as a key skill to be not only promoted in official rhetoric but

also through tangible initiatives such as mother tongue instruction, the ProMES project, Barcelona Growth, and Barcelona Activa projects.

4. The different plans and documents of the City of London highlight the importance of linguistic diversity and depict it as one of London's greatest assets but fail to translate it into concrete and tangible actions. It is only at grassroots level that we can find a whole range of initiatives.

As highlighted in this paper, the current world scenario points at an increasing concentration of people in cities. The population is more mobile, diverse and heterogeneous than ever and managing this diversity will become one of the priorities over the years to come. The current economic and social changes taking place around the world are producing an unprecedented change in trends and patterns pointing in this direction. At the core of these changes, we find cities, the centres of economic activity. In the urban knowledge society, linguistic diversity, far from being a problem, can become a city's greatest asset. A source of prosperity, growth, progress, cohesion and inclusion with multiple personal, social and economic benefits: personal, because while it is true that learning the host society's language can be a source of integration and participation, immigrants speak languages that can be extremely useful in maintaining personal bonds; social, because they enrich the host society as a whole, making it a more diverse, plural society; and economic, because the languages spoken by migrants are those of emerging countries, which can prove extremely useful when making business in the countries of origin. Linguistic diversity, often portrayed as a problematic Tower of Babel, can be a society's greatest asset.

References

1. Berliant, M. & Fujita, M. (2003). *Knowledge creation as a square dance on the Hilbert Cube*, Washington University, Department of Economics, Mimeo.

2. Castles, S. and Miller, M. J. (2003). The Age of Migration (Third Edition), Macmillan, London.

3. Favell, A. (2001). Philosophies of Integration: Immigration and the Idea of Citizenship in France and Britain (2nd ed.). London: Macmillan and New York: St. Martin's Press.

4. Freeman, O. Business partnerships with migrants' countries of origin: sharing the diversity advantage.

5. Kraus, P. (2011). "The Multilingual City. The Cases of Helsinki and Barcelona". *The Nordic Journal of Migration Studies* 1(1).

6. Landry, C. & Wood, P. (2008). The Intercultural City: Planning for Diversity Advantage. Earthscan, London.

7. London city Council. (2007). The Impact of Recent Migration on the London economy. London School of Economics.

8. Pyke, J. (2005). *Productive Diversity: Which Companies Are Active and Why?* Victoria University, Phd Thesis.

9. Sassen, S. (2001). *The Global City*, 2nd ed. Princeton University Press, Princeton.

10. Vertovec, S. (2007a). 'Super-diversity and its implications'. In: *Ethnic and Racial Studies* 29(6): 1024-54.

Web References

- International Migration Report 2013: http://www.un.org/en/development/desa/population/publications/pdf/migration/migrationreport2013/Full_Document_final.pdf#zoom=100.

- European Commission. European Agenda for the Integration of Third-Country Nationals. Communication from the Commission: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0455:FIN:EN:PDF.

- Migration and the United Nations Post-2015 Development Agenda: http://publications.iom.int/bookstore/free/Migration_and_the_UN_Post2015_Agenda.pdf.

- UN Habitat: http://www.unhabitat.org/pmss/listItemDetails.aspx?publicationID=3387.

- Council of Europe's Intercultural Cities project: http://www.coe.int/t/dg4/cultureheritage/culture/Cities/ICCOutcomes_en.pdf.

- Evidence of the Economic and Social Advantages of Intercultural Cities Approach: http://www.coe.int/t/dg4/cultureheritage/culture/cities/research/literature%20review.pdf.

- Languages of Toronto. Toronto City Council: http://www.toronto.ca/demographics/pdf/language_2011_backgrounder.pdf.

- Language and Education in New York. Migration Policy Institute: http://www.migrationinformation.org/datahub/state2.cfm?ID=NY.

- University of Manchester. The Languages of Manchester: http://mlm.humanities.manchester.ac.uk/aboutus.html.

- The Independent London: Multilingual Capital of the World: http://www.independent.co.uk/news/london-multilingual-capital-of-the-world-1083812.html.

- Catalan Institute for Statistics (IDESCAT): www.idescat.cat.

- UK Census Analysis. Languages in England and Wales 2011: http://www.ons.gov.uk/ons/rel/census/2011-census-analysis/language-in-england-and-wales-2011/index.html.

- CILT, the National Centre for Languages. ELAN: The Effects on the European Economy of Shortages of Foreign Language Skills in Enterprise, 2006: http://ec.europa.eu/languages/policy/strategic-framework/documents/elan_en.pdf.

- Migrants Contribute. International Organization for Migration: http://migrantscontribute.com/.

- Barcelona 2012–2015 Immigration Plan: http://www.bcn.cat/novaciutadania/pdf/pla_immigracio/pla_immigracio_en.pdf.

- The London Plan: https://www.london.gov.uk/priorities/planning/london-plan.

- London School of Economics. LUCIDE project: http://www.urbanlanguages.eu/events.

Vassili RIVRON

Researcher and Social Science Coordinator for Metroscope,
National Institution for Informatics and Automation,
University of Caen Lower Normandy
(Caen, France)

Social Media and Linguistic Affirmation in Central Africa. Between Cultural Objectification and Cultural Mutation

Abstract

The success of social media among Africans and African diasporas have led to the creation of "Facebook groups" identified as ethnic groups. These networks can gather, among the five groups included in this study, up to 8,000 participants each. These spaces of community exchanges allow interesting observations on the preservation of linguistic diversity in the era of digital globalization, and on the mutations of ethnicity.

On one hand, we are witnessing the "spontaneous" encoding of languages that were not usually written, contributing to their current writing use, to the transmission of this competence and literary heritage, to its unification and homogenization.

On the other hand, these new linguistic registers cause profound changes in the status of these languages and in the organization of the associated cultural groups: the creation of a public space (which partially excludes other national languages) where diaspora members play a central role; the weaving of supra-national links with former parent groups in neighboring countries (reinvestment of neglected ethnonyms, rewriting of genealogies, reaffirmation of unifying origin myths); the projects of physical meetings, publishing policies, cultural festivals, supra-national political parties...

Introduction

The Eton language, spoken mainly in central and southern Cameroon, belongs to the numerous Bantu languages that spread from central to South Africa. Cameroon has 2 official languages (French and English), and around 200 native languages of which very few are written and taught, as school options. The Eton language is spoken by around 250,000 people. This language doesn't have a properly codified grammar, syntax and orthography, in the sense of specific and usual rules established to write it, and before

social media, there were few contexts for scribal practices and even rarer were written publications.

In a programmatic paper [Rivron 2012] we had presented a sort of collaborative process of scripture codification and practice of this mother tongue, on social media and forums that emerged around 2009. At that time, several studies identified similar phenomena for the Amazighs in North Africa [Azizi 2010], or the Hmongs in China [Mayhoua 2010], for example, showing how Web 2.0 resources could contribute to consolidate cultural communities. Since then, there has been a significant evolution of this field in Central Africa. As the local connectivity and the movement of cultural affirmation grow on the Internet, we now have new kinds of "cultural" or "ethnic Facebook groups" from that region, several of them consisting up to 8,000 members. There has been a scale effect by which the most popular "cultural groups" now join several inter-comprehension ethnic groups, upon the linguistic category of "Fang-Bulu-Beti" and across Cameroon, Gabon and Equatorial Guinea state frontiers.

The arising of current native literacy on Facebook is associated with a process of "patrimonialization", a formal construction of cultural heritage: concentration, unification and codification of cultural capital [Bourdieu 1994]. This patrimonialization shows typical but also very specific characteristics, in which we have the illustration of a hypothesis that was formulated by Renato Ortiz [1995]: "The fixation of traditions always happens in a modernization process". With fixation of traditions I understand "objectification" of culture in the sense of Jack Goody [1979]: the materialization of cultural traits through the production of communication supports that permits a systematization and a distant transmission, in time and space. And as Eric Guichard [2003] stresses, informatics and the Internet are based on scripture (from code to contents) and should also be analyzed as "intellectual techniques" which have social effects.

The present approach of linguistic and cultural affirmation is articulated to Jean-Loup Amselle and Elikia Mbokolo's asserts in *Au coeur de l'Ethnie* [1999], where they develop a non-static comprehension of culture and ethnicity: these are relational constructions (isolated cultures are exceptions); those are dynamic, not "cold societies" as Claude Levi-Strauss would have said; the cultural categories and practices are polysemous and contextual, not essentialists... So, we'll try to analyze here how the preservation and the transmission of languages, through their written codification and electronic sociability, take part of a transformation of how groups think and reconfigure themselves.

Eton Scripture in Facebook Groups

In 2009, I was very surprised at discovering several attempts of recomposing ethnic communities on one of the most famous web co-optative social network (or social media). These practices happened in a context in which Eton was sometimes affirmed by its speakers to be in a threat of getting disappeared or, more often, of being degraded by a lack of practice and a trend to mix with other languages like Ewondo. On the other hand, this mother-tongue is not registered in the UNESCO Atlas of threatened languages [Moseley 2010], and we could think that the claims of being in danger were strategies of affirmation of cultural value in a "rhetoric of loss" [Gonçalves 1996] that is common to the patrimonialization processes. In the absence of statistical documentation, we can daily make the confirmation of that intergenerational lack of transmission in the urban Eton communities, a lack mainly due to the processes of migration, urbanization, upward social mobility, globalization of cultural industries and creolization between neighboring inter-comprehensive thongs.

This mother tongue hasn't been specifically codified for a usual writing and until then, it was only written in exceptional conditions: by linguists, ethnographers, folklorists, artists, and probably in a few intimacy scriptures. The current written practice of Eton in a "public space" began sporadically through electronic networks. We first saw it in 2004 on static web sites and web forums (connivance and interjections only), before the Facebook groups saw the appearance of real written conversations. At the same time, this process showed a non-exclusive use of written Eton, beside French and English that occupy the main textual space.

Far from the futile conversations that we could imagine on these social networks, deep discussions emerged on how to speak correctly (up to bind the generational and geographical gaps), and how to spell this language. Further appeared redundant debates about etymology, rituals, history, genealogy, political structures, regional news...

Several times, we saw posts making reference to a French-Eton dictionary (which is in reality a PhD word that includes a lexicon), and a grammar of Eton (written in English) by the same author: the Belgian linguist Mark Van Der Veld [2003, 2008]. If these texts unify and systematize the spelling of the language, the author himself (in an interview with the author, January 12th, 2011) recognizes the need to develop practical tools for a usual scripture of this language. These two documents use complex linguistic considerations, with phonetic symbols that are not easily understandable or even available on common keyboards. So the reference to these two documents had probably a valorization impact but may practically be used only by the most erudite

members of these communities: within the thousands of members of these groups, we only have seen very rare people using its complex phonetic scripture (5 or 6 persons).

The short texts and conversations in Eton are mainly written in alphabetic characters, taking inspiration from the Ewondo scripture (a close language from Yaoundé, that was codified by missionaries and is eventually taught at school as an option). They often make total abstraction of the tonal aspects of the language, proceeding to a graphic reduction, in the sense of Jack Goody (the transcription reduces the richness of oral and contextual communication), but also in a sense pointed by Mark Van de Veld himself in our interview: the missionaries "heard" less tones than used in oral situations, having a lexical loss as a consequence. The emergence of this scribal practice, mainly in urban contexts, explains also the many borrowing from the Ewondo vocabulary, and the need to read at loud voice or with the lips, sometimes several times, to comprehend the text.

Effects of the Scripture: Objectification, Patrimonialization and Graphic Reduction

These Cameroonian "ethnic" or "cultural" groups on Facebook do not limit themselves to trying to compensate the lack of intergenerational transmission in the context of urbanization, migration and upward social mobility. They also make an appropriation of the generic resources offered by the Facebook platform, in order to integrate or project traditional sociability codes, procedures and rituals. The insistence on formal presentation for the new comers by passing through the "house of presentations" in one of those groups, or the existence of topics dedicated to marriage transactions or rumor spreading are intents to integrate electronic sociability resources into cultural patterns.

As well as the impact of "materializing" the speech by scripture, other "objectification" processes occur to preserve and transmit cultural traditions, producing at the same time new senses for the same linguistic categories that Ortiz [1995] mentioned when talking about the modernization implied in every attempt to fix materially the traditions.

One of those is the production of collaborative and public archives, when historical documents such as photos, videos and texts are progressively compiled into the news feed or into attached documents. It truly participates to the patrimonialization process, proceeding to a new kind of concentration, unification and capitalization of information of all kinds, that was initially spread in personal archives, into a same graphic support. Moreover it valorizes

visual symbols of ethnic belonging that had been partially erased, mainly by Christian conversion and colonial administration.

Another modernization effect, that can be noticed – even if it's implications do not come exclusively from these practices – is the territorialization of the ethnic imaginary. Computer resources and Internet spreading are followed by a vast cartographic production that is also present in the observed material. As Paul Bohannan [1963] showed, African ethnic groups are not always identified to a certain territory, and their settlement and territorialization result mainly from colonial and independent administration.

The general effect of these Facebook ethnic sociability and patrimonialization is the production of cultural pride: a positive perception of traditional belonging in a modern and globalized life where social mobility tends to depreciate mother tongue and ethnicity as archaisms.

Scale Effects and Revitalization of the "Ekang" Ethnonym

Quickly after the creation of two "Eton" groups on Facebook, which remained relatively small (up to 1,500 members), several Facebook groups were created with a much broader scope: the "Beti" and the "Fang-Bulu-Beti". These appeal to a larger cultural definition of cultural group: the inter-comprehensive space between South Cameroon, Northern Gabon and Equatorial Guinea, that had initially been identified separately by linguists. Several groups have followed that broader concept and each of them can gather now from 3,000 to 8,000 members.

In Figure 1, we have concatenated[102] 15 days (in November 2013) of interactions inside and between four "meta ethnic groups" of this kind, including another group elaborated specifically for the Cameroon Diaspora (A). The nods correspond to posts made by individuals, and the links indicate shares, likes and comments by other members (interactions). The legend indicates the percentage of each group interactions or intersection between groups' interactions in the corpus. For the actual black and white publication of this figure, we had to add the approximate perimeters of the groups, and the interactions between the groups (multiple belonging of individuals and content circulation between different groups) do not appear clearly.

[102] *With the great help of Simon Charneau, engineer at INRIA.*
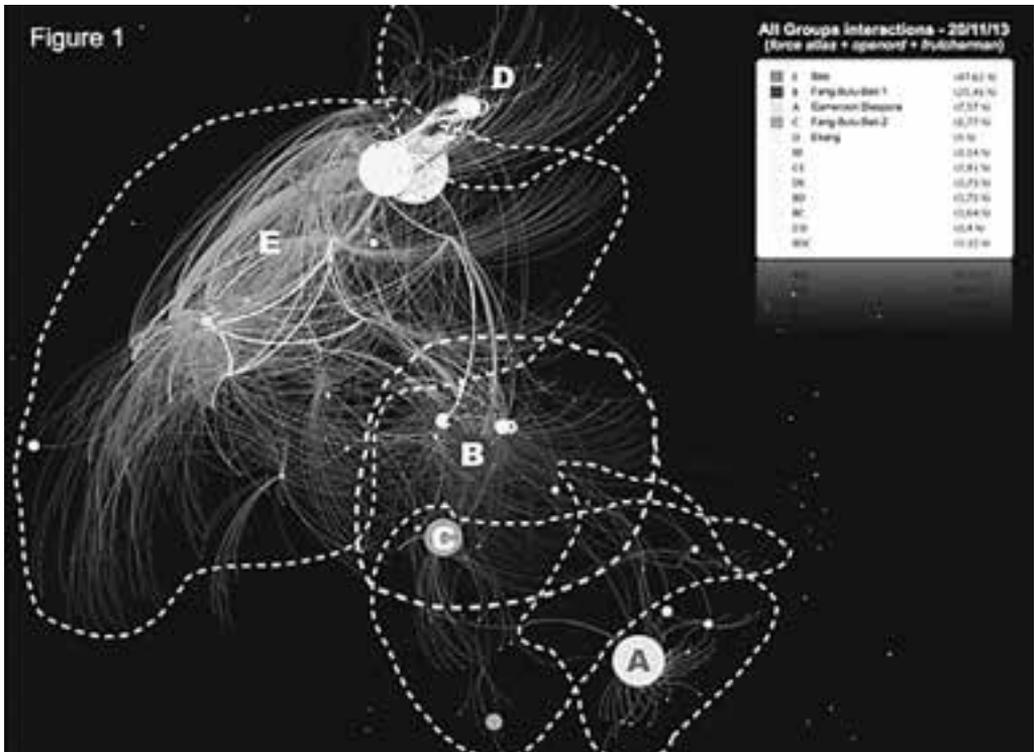
Figure 1. *Meta ehnic groups interactions*

The global topology of these networks reveals very different kinds of interactions and organizations. Groups A, B and C have a very centralized structure of interactions. For example, B has a main contributor/moderator that posts very regularly (almost every day) on cultural or linguistic topics through explicit questions, and the members react with answers also very regularly, but rarely contribute with their own posts. This group looks like a school or church interaction: a main actor driving all the publication initiative, and the students or believers follow in a very disciplined way. On the contrary, groups D and E have a very distributed structure corresponding to another publication dynamic: there are many contributors with posts on cultural, media, politic or commercial topics, and the interactions around these go even outside the group (for example here, the share of several contents from E, into D, and important contributors that work actively between these groups without centralizing the activity of the groups). The analysis of these groups' memberships also reveals an important interpenetration between these groups: some have up to 20% of their members belonging to one or more groups of this category.

If we focus on the two "Fang-Bulu-Beti" groups (B and C), they show a specific activity around patrimonialization, mother thongs scripture and cultural identity production. And they simultaneously led to the rise of a different ethnonym, "Ekang" or "Ekañ", that became a topic of discussion and also the name of the further Facebook groups corresponding to the Fang-Bulu-Beti linguistic cluster. Working on Eton and Ewondo fields since the end of the 1990's, we had never noticed that cultural category, even if we later confirmed its presence in cosmogony epics sang on the Mvet. A quick research through "Google Trends", that counts the occurrences of keywords queries, also showed that "Ekang" wasn't used on the web before 2011.

So we assist through these Facebook ethnic activities to the restoration or the reactivation of this category to designate a new kind of meta-ethnic group whose existence was limited before that, to the linguistic association of the Fang, the Bulu and the Beti. What's noticeable about it, is that the cultural investigation initiated in these groups on linguistic proximities and etymologic considerations, finally led to the unification of common genealogies, to the formulation of common origin myths and the identification of common cultural patterns. The discourses about this Ekang reunification claim the revitalization of cultural and kinship relations that had been forgotten with colonization and state-building process.

Political Stakes of Community and Linguistic Process

The first approach could make us believe in a perfectly virtual (online), spontaneous and horizontal process, corresponding to the ideology of Internet promoters and distant education. But we soon discover the central role of specific and dominant components of these different ethnic groups. Processes on social media are not only factors, but also indicators of the actual dynamics of social, economic and political interdependences and organization.

A demographic approach of the Eton and the Ekang groups on Facebook – even if very basic because of technical limitations as well as due to the raw quality of the information on Facebook profiles – show that this ethnic affirmation and construction is driven mainly by urban, cosmopolite and diaspora scholars. And these characteristics are even more predominant if we consider only the main activists of these groups, which write mainly from France, the United States, Canada, Belgium and Germany, when they are not at Yaound or Libreville.

This is not surprising if we consider the still weak Internet and Facebook penetration in Cameroon, Gabon and Equatorial Guinea, compared to other countries in Europe or America, and the inequalities existing towards written and computer techniques. The composition of the Ekang Facebook activists

is neither surprising, if we consider the political stakes that are involved in this subsequent unification project, that even resulted on physical meetings and projects of cultural festivals (defended by a culture promoter and book editor). The arising of a Fang-Bulu-Beti public space, in the sense of Habermas, questioning political boundaries erected by colonization and state building, also ended to the hypothetic project of a supranational political party for the CEMAC (Economic and Monetary Community of Central Africa).

Conclusion

From the initial observations about the Eton social media, to the actual Ekang dynamics on the web, we can clearly see that the Internet and Social Media are effective tools for preserving and developing mother tongues. However, several linguistic stakes remain: the development of written conversations between inter-comprehensive languages is clearly a factor of language valorization. But should we consider it as a threat for the scripture of languages like Eton, when already codified and taught scriptures of other languages are the predominant resources? And would the unification process of these mutual understanding tongues result into a new kind of *lingua franca* or *standard scripture?*

The hypothesis of this standardization, that would include a much broader population, is perhaps the condition for the development of electronic resources in these languages (software translation, online publication, electronic dictionaries or translators). But it is also a threat to the original mother tongue and to linguistic diversity...

And even, if these electronic sociabilities seem to represent a good hope for the vitality and perpetration of mother tongues from Central Africa, it will hardly become by itself a systematized and durable solution for codifying non-written languages and teach them to non-speakers. There is a need of organizing and systematizing that won't happen into these social media, but into authorized and organized institutions and intermediaries. And this would be a totally different and a more classical process, where social hierarchies and state building dynamics would reappear clearly.

References

1. Amselle, J.-L., Mbokolo, E. (dir.) (1999). *Au cœur de l'ethnie*, Paris, La Découverte.

2. Azizi, S. (2010). "Les Idaw Facebook. Typologie de groups Amazighs sur un réseau social virtuel". In: *La culture amazighe. Réflexions et pratiques anthropologiques.*

3. Bohannan, P. (1963). "Land, tenure and land-tenure". In: Biebuyck, D. *African agrarian systems*, Oxford University Press, Oxford, p. 106 et suiv.

4. Bourdieu, P. (1994). "Esprits d'état: Genèse et structure du champ bureaucratique". In: *Raisons Pratiques*, Seuil, Paris, pp. 99-135.

5. Gonçalves, J. (1996). *A retórica da perda. Os discursos do patrimônio Cultural no Brasil*, Ed. UFRJ/MinC-IPHAN, Rio de Janeiro.

6. Goody, J. (1979). *La raison Graphique,* Éditions de Minuit, 274 p.

7. Guichard, E. (2003). "Does the 'Digital Divide' Exist?". In: *Globalization and its new divides: malcontents, recipes, and reform* (dir. Paul van Seters, Bas de Gaay Fortman & Arie de Ruijter), Dutch University Press, Amsterdam.

8. Guyer, J. I. (2000). "La tradition de l'invention en Afrique équatoriale", Politique africaine, n°79, octobre, pp.101-139.

9. Mayhoua, M. (2010). "Diversité culturelle et usages du web. Les pratiques communautaires à travers le forum "hmong"". *Réseaux*, 1 (n°159), p. 199-218.

10. Moseley, C. (ed.) (2010). Atlas of the World's Languages in Danger, 3rd ed. Paris, UNESCO Publishing.

11. Ortiz, R. (1995). *A moderna tradição brasileira – cultura brasileira e indústria cultural*, ed. Brasiliense, São Paulo.

12. Rivron, V. (2012). "The use of Facebook by the Eton of Cameroon", in *Net.lang: Towards the Multilingual Cyberspace,* C&F éditions, pp. 160-166.

13. Sayad, A. (1985). "Du message oral au message sur cassette: la communication avec l'absent", Actes de la recherche en sciences sociales, n°59, pp. 61-72.

14. Van de Velde, M. (2006). *A description of Eton: phonology, morphology, basic syntax and lexicon,* thèse de doctorat.

15. Van de Velde, M. (2008). A grammar of Eton, Mouton de Gruyter/ Department of Linguistics, University of Leuven, 432 p.

Virach SORNLERTLAMVANICH

Advisor to Executive Director,
Technology Promotion Association (Thailand – Japan)
(Bangkok, Thailand)

Understanding Social Movement by Keyword Tracking in Social Media

Introduction

Social media is massive communication data for understanding the social behavior as well as sensing network is massive monitoring data for observing the global environment. Both are generated data reflecting the real-time current situation of society and environment. In the rapidly changing modern world, it is necessary to understand the situation and make a suitable response timely. The effect of happening or disaster nowadays has a trend to cause tremendous and pervasive damages. Since the Great Hanshin earthquake in 1995, the Indian Ocean earthquake and tsunami in 2004, the Illinois hurricane Katrina in 2005, Arab spring (a series of anti-government protests in 2011 uprising in Tunisia spread out to Yemen, Egypt, Syria, Libya and most of Arab countries), the Tohoku earthquake and tsunami in 2011, the Occupy Wall Street movement in 2011 and until the recent Thailand coup d'etat in 2014, it is wondered whether we can learn something about these historical events. Focusing on social happenings, it is efficient enough to collect social media data from widely used applications such as Facebook, Twitter, WhatsApp, Line, or WeChat. Social media are actively used in most of the recent cases. If we ever view them in a proper dimension it is no doubt that we can somehow forecast, prevent, avoid happenings by warning or influencing the communities to relief the disaster or the undesirable social situation development. In reality, social media data are vast, noisy, distributed, unstructured, and dynamic. [Gundecha and Liu 2012]

To study the evolution of social behavior on a happening, we analyze the time series of tweets related to the topic of the recent Thailand coup d'etat in 2014. According to a 2013 survey[103], there are 12 million twitter users in Thailand with 200,000 active users/day. This means that if we can screen for the related tweets we can observe the movement of the community tie-up.

---

[103] *http://www.techinasia.com/thailand-18-million-social-media-users-in-2013/.*

In our experiment, we estimate the topic related keywords from the target document that we can simply collect from the Internet news. A tweet is a short 140-character text, which is more likely to be a conversational text comparing to a written document, which is some kind of political news or a review. There is a difference in the extracted keywords. We therefore apply a technique in GETA (Generic Engine for Transposable Association) called WAM (Word Article Matrix) [Murakami et al. 2004] to expand the set of keywords reflecting the nature of the text from Twitter.

The transition of a word cloud in a time series can express the social interest at the moment. From a set of related tweets, we extract keywords and express them in a word cloud manner. We then put the word cloud on the time series to create a timeline word cloud. The word cloud each moment expresses the social interest, which significantly changes at the time of happening.

Keyword Expansion

WAM (Word Article Matrix) is a table of weighted relation between a document and keywords. Keywords in the document are counted to fill in the table.
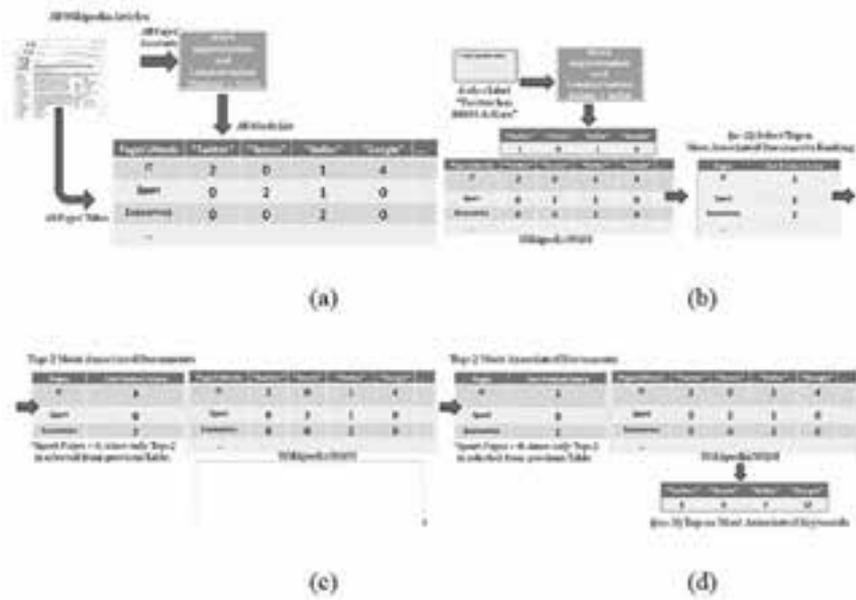

**Figure 1.** *WAM and keyword expansion*

WAM is created in Figure 1(a) when the input documents are word segmented (in case of non-segmented language such as Thai) or lemmatized, and the corresponding keywords are counted. The matrix is used to operate dot matrix

with the input of a training set of tweets shown in Figure 1(b). As a result, a table of the most associated documents to the training set is obtained. The ranked documents can be cut off by setting up a threshold for the associated value as shown in Figure 1(c). With another dot matrix in Figure 1(d) the expanded associated keyword can be obtained with the weight. By training through the set of targeted tweets, associated keywords in the target domain can be created. Now we can rank the keyword by its associated weight to retrieve the topic related tweets from Twitter.

Timeline Word Cloud

Figure 2(a) shows the process of creating a Twitter word cloud. A set of topic related documents are collected to create a WAM. The WAM is used to expand the keyword from the initial set of tweets. The iterative operation in expanding the keyword allows us to query Twitter for better coverage of the tweets. Under the constraint of 100 tweets/query and 7 days search back, we repeatedly issue the query using Twitter search API with the set of keywords. As a result 339,148 tweets centering on the date of coup d'etat on May 22, 2014 are collected. On each day the word cloud is generated to compare on hourly basis.

Investigating the happening that the National Peace Keeping Committee seized power on May 22, 2014 at 4.30 p.m., Figure 2(b) shows the transition of a word cloud around the target time. Significantly the word "coup d'etat" occurs every hour as the most focusing topic. Before the time of announcement, it is obvious that the Twitter community is already alert to the possibility of coup d'etat. The density of the keyword increases significantly along the climax time. The timeline word cloud explicitly shows the critical change point of the happening. Strategic planning can be considered to handle the happening by observing the effective timeline word cloud.
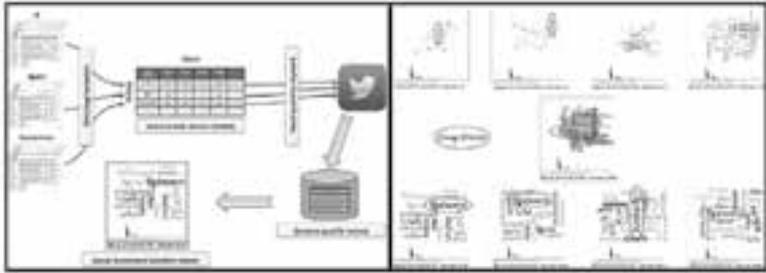
(a)   (b)

Figure 2. *Word cloud and timeline word cloud*

Conclusion

A timeline word cloud is an effective instrument to monitor the social behavior since the community tie-up of the social media users is reliable. In the modern Internet use, the growth of social media as well as the sensing network is not ignorable. Understanding the movement of interest in the social media community can be beneficial in the process of strategic planning or decision-making. In the coming future, spatial-temporal information can be inclusively considered to create a wider dimension in monitoring the movement and happenings can be understood in a more precise manner.

References

1. Gundecha, P. and Liu, H. (2012). Mining social media: a brief introduction. Tutorials in Operations Research, Informs, 1(4).

2. Murakami, T., Hu, Z., Nishioka, S., Takano, A., and Takeichi, M. (2004). An Algebraic Interface for GETA Search Engine. Proceedings of the Programme and Programming Language Workshop, Japan.

SECTION 3.

PRESERVATION OF LINGUISTIC AND CULTURAL DIVERSITY IN CYBERSPACE: NATIONAL VISION AND EXPERIENCE

Panchanan MOHANTY

Professor, Centre for Applied Linguistics & Translation Studies,
Centre for Endangered Languages & Mother Tongue Studies,
University of Hyderabad
(Hyderabad, India)

Conservation of Linguistic Diversity: The Indian Experience

Abstract

Though conservation of biodiversity has been a buzz word for a long time throughout the world, conservation of linguistic diversity has gained momentum only recently. A cursory look on the languages spoken in this world reveals that 33% of them are spoken in Asia. If we take the Indian scenario into consideration, according to the 2001 Census, 96.56% of Indian citizens speak the few scheduled languages whereas only 3.44% speak hundreds of non-scheduled languages and most of these are in the endangered category. It means if these 3.44% of Indians by chance shift to some other dominant language, India will no more boast of being one of the most linguistically diverse countries in the world. Now the question is: why is the linguistic diversity so important? UNSECO has emphasized the point that cultural diversity is an important driving force of development and linguistic diversity is an important factor for cultural diversity. It also has stated that cultural diversity, linguistic diversity and biological diversity are interdependent. Needless to mention that only if the above mentioned hundreds of minor and endangered languages spoken by the 3.44% of Indians are conserved, we will be able to maintain our linguistic diversity. Therefore, conservation of endangered languages is very important for a country like India. What is significant here is that when most of the indigenous languages in the developed countries like the U.S.A., Europe, and Australia have become extinct, they have survived very well in India for millennia. But there is a change in the situation in recent times and now we find that a sizeable number of these minor and indigenous language speakers

are switching over to a major and dominant language. Keeping these in view, I want to discuss in this paper how and why these languages could survive for so long in India along with the causes of their sudden decay in the globalised modern environment.

1. Introduction

Though biodiversity and linguistic diversity are very much interdependent, most people including the elites and those pressure groups whose views matter do not realise it. That is why news items regarding the efforts to preserve languages are extremely rare in the Indian news media. What is shocking is that about one third of the total population of the world's languages has not been documented yet. It means there is no record of more than 2000 languages. If by chance, these languages die before they are documented, we will lose an enormous amount of indigenous knowledge without even knowing what existed on this earth that we have lost. This is certainly not commensurate with what the world's intellectuals are interested in.

The Summer Institute of Linguistics Survey published in 1999 presents the following figures about the status of a number of languages under the threat of extinction: There were 51 languages with one speaker each and 28 of them were in Australia. About 500 languages had less than 100 speakers, 1500 languages less than 1000 speakers, 3000 languages less than 10,000 speakers, and 5,000 languages less than 100,000 speakers. If we tabulate these data, we will see that a staggering 96% of the languages are spoken by a meagre 4% of the population and a meagre 4% of the languages are spoken by a staggering 96% of the population. Even though this situation is visibly catastrophic, it has not been able to draw attention of the people who matter in this world. On the basis of two judgements from the Foundation for Endangered Languages that about 50% of the world's languages will face extinction in the next 100 years, Crystal [2000: 19] has calculated that "To meet that time frame, at least one language must die, on average, every two weeks or so. This cannot be very far from the truth." There is a still more alarming prediction that "...up to 90% of the world's languages may well be replaced by dominant languages by the end of the 21st century, which would reduce the present number of almost 7,000 languages to less than 700." [Brenzinger & de Graaf n.d.:1]. But the irony is that, unlike the issue of preservation of biodiversity, hardly any serious steps are being taken at the national or international levels for protection of these fast vanishing voices. A few volunteer groups have been trying to make their voice heard by the world's citizens, but it is not loud enough and hence the outcome is far from being adequate.

There is near-complete silence with reference to the situation in India where plurality is practised and variation is expressed in the day-to-day life. This has naturally led to discrimination among the languages of India in spite of the constitutional safeguards. Due to economic and political reasons some languages are perceived to be more than equal by other speech communities, and that is why the domains of the latter languages are slowly shrinking giving way to the former to occupy those spaces. In this paper, I will discuss some of these issues along with suggestions for empowerment and revitalisation of the endangered languages with reference to the Indian context.

2. Loss of Indigenous Languages in the Developed World

The United States of America spoke some 250 to 350 indigenous languages before the arrival of the European immigrants more than two centuries ago [Crawford 1999]. Linguistic and cultural diversity was tolerated in that country for about a century and after that an English-only attitude spread and occupied both public and private spaces. For instance, an English-only instruction was mandated in California in 1855. Then, the Bureau of Indian Affairs started implementing its language suppression policies ruthlessly around the 1880s [Ibid.]. As a result, English bossed over the innocent minor languages and, over a period of time, English monolingual education became the mainstream. The Americanisation Department of the United States Bureau of Education passed a resolution in 1919 recommending that education in all schools, i.e. both public and private, be imparted in English including the elementary classes [Garcia 1992].

In spite of the 1964 Civil Rights Act that regulated against discrimination on the basis of colour, race and origin, intervention of the judiciary as well as efforts of various activist groups, English remains the dominant medium of instruction in the U.S. educational system. Regarding intervention of the judiciary, the 1970 lawsuit by some non-English speaking Chinese students against the San Francisco School District is considered a landmark and known as the 'Lau vs. Nichols case'. Though it was rejected by the federal district court and another court of appeals, the Supreme Court admitted it in 1974 and ruled: "There is no equality of treatment merely by providing students with the same facilities, textbooks, teachers and curriculum; for students who do not understand English are effectively foreclosed from any meaningful education." (quoted in [Baker 2001: 186]). As a result, a compensatory "poverty programme" was introduced which encouraged a few teachers to use the "home languages" in classroom while imparting education in English. The name "poverty programme" indicates the government's attitude towards the people who are supposed to go through it. Even President Ronald Regan

was quoted to have said the following in *The New York Times* of 3 March 1981 with reference to educating the linguistic minorities in their languages: "It is absolutely wrong and against the American concept to have a bilingual education programme that is now openly, admittedly, dedicated to preserving their native language and never getting them adequate English so they can go out into the job market." [Ibid: 187-8] Hence, the famous joke I have heard from my American friends: "If you speak many languages, you are multilingual. If you speak two languages, you are bilingual. If you speak one language, you are American." Therefore, it is no wonder that 155 out of the 175 native American languages are in the endangered category [Krauss 1992].

Let us now have a bird's eye-view on the situation in Europe. The European Union declared the year 2001 as the European Year of Languages and the desired goal was to lay the foundation for a multilingual Europe: "All those leaving compulsory education should be able to communicate in at least two European languages in addition to their mother tongue and then be able to build on that knowledge for the rest of their lives."[2] But the ground realities seem contradictory. Many European countries believe that "...knowledge of the heritage language is something that is unnecessary at best and detrimental to integration into the dominant society at worst." [Schmid 2002: 355] This is clearly reflected in withdrawal of funding for various heritage language teaching programmes in European public schools.

The situation in Australia is no better either. After a detailed analysis of the Australian situation, Fishman [1991] has come to a grim conclusion that these aboriginal languages do not have any long-term survival prospects. Whatever may be the propaganda regarding protection and preservation of the "small" languages, all these facts uncover the stark realities prevailing in the developed world.

3. The Indian Scenario

Of late, scholars have started presenting case studies on language loss and language shift in different linguistics seminars and conferences in India. It is undoubtedly a good sign, because it is indicative of increase in the level of awareness regarding language loss/shift/death. But there are two questions that demand attention:

1. Is the Indian situation comparable to those prevalent in the U.S., Europe and Australia?

2. Is it enough to discuss the problems only? If not, what are the remedies?

Let us discuss these questions in some detail.

A close look at the Indian situation makes it evident that it is different from the situation in the developed countries. As we have seen above, monolingualism is the fundament on which the linguistic structures have been built in the western world whereas multilingualism is the very basis of the Indian society. Most of the *adivasi* or primitive languages have diminished in the U.S. and Australia within hardly two centuries whereas hundreds of languages have survived on the Indian soil through millennia. Indian languages have learnt to co-exist though they have converged with each other and at the same time maintained their individual identities. This has been demonstrated in a fascinating study by Gumperz and Wilson [1971] on the linguistic situation in an Indian village named Kupwar. The three languages spoken in this village are Kannada, a Dravidian language, and Urdu and Marathi that belong to the Indo-Aryan family. Bilingualism and multilingualism in this village are widespread, and it is common to see people switching back and forth between at least two languages. This has made the grammatical structures of these three languages so similar that a word-for-word rendering among them is possible. Finegan and Besnier [1989: 386] have called this phenomenon "Having your cake and eating it too." Though Myers-Scotton [1992: 33] has argued that "... code-switching is involved in language death." and "...at least some instances of language death may involve the pervasive addition or substitution of the grammar of another language in the code-switching situation", the Kupwar example cited above does not substantiate this argument. Since she has elaborated her views on the basis of African languages, it can be claimed that the linguistic situation of India is different not only from that of the Western world, but also from the African one.

It is, of course, true that many languages known to us have died over a period of time, and such a natural death is inevitable. Some major and powerful languages like Sanskrit and various Prakrits that received support from rulers, religious leaders and masses are included in it. But so far as minor languages are concerned, Lo Bianco and Rhydwen [2001: 394] have proposed two types of loss: "...an abrupt dislocative and extreme form and a slower, generational, attrition. The former often results in the total disruption of all transmission (and of any later re-learning prospects) of the language, while the second can, at best, retain within the living memory of speakers sufficient language resources on which to base a revival or renewal activity. The former may be called language loss by *rupture* and the latter language loss by *attrition*." Though they make this distinction with reference to the indigenous languages of Australia, it is applicable to the loss of minor languages in general. Interestingly, "language loss by rupture" is typically a western phenomenon and it has hardly taken

place in India. But the second type of loss, i.e. "language loss by attrition" is a common feature across the Indian subcontinent.

But it is interesting to note when the western world has been quite enthusiastic and active in protecting small and endangered languages. In his "General introduction" to the *Encyclopedia of the World's Endangered Languages,* Christopher Mosley [2007: x] states: "In the past decade leading up to the publication of this encyclopedia there have been various initiatives to foster awareness of the accelerating rate of the loss of languages. The UNESCO Red Book, which appeared in 1993, was a pioneering effort in this direction. Then in 1995 the University of Tokyo set up a Clearing House for endangered languages, the emphasis being on recording newly discovered instances of disappearing languages rather than taking action to preserve them. There swiftly followed the creation of "activist" groups on both sides of the Atlantic in 1995: in the USA, the Endangered Languages Fund (ELF), and in Britain, the Foundation for Endangered Languages (FEL). These bodies have taken an active part in the actual preservation of endangered languages, by acting as charitable grant-giving bodies which make awards to scholars who are doing valuable investigative work; they stipulate that the published results of the work undertaken shall benefit the community concerned. The prestige of the study of endangered languages was further enhanced in 2002 with the creation of the first university chair in the subject by the Rausing Foundation at the School of Oriental and African Studies in London. This foundation, too, is engaged in giving grants for research projects that it deems will assist in language recovery." It is unfortunate that any such steps are so far unheard of in India. Except at the individual level, language endangerment is not an independent subject of study in most Indian universities. Though the linguistic situation in India is inextricably complex, the language planning activities certainly do not match with that complexity and, therefore, have been superficial to a great extent. Schiffman [1996] has proposed the concept of "linguistic culture" that greatly influences the prospects of a country's language policy. He has argued that unless we understand the covert linguistic culture of a country that consists of the belief systems, ideas and attitudes about languages, we cannot achieve success in implementing the overt language policies. I am not aware of any study conducted to discover the covert linguistic culture of India. So, it is natural for the overt language policies to lay hidden in the files of the government offices.

We know that the Indian society is in an unprecedented flux and transition and a large number of Indians belonging to different strata have been making efforts to join the national/international mainstream which has been radically affected the global economic changes. As a result, the good old joint

family structure has broken down and nuclear families are the trend now. All these people live in "houses" and the concept of "home" has become a thing of the past. They are uprooted from the "home-land" and have left the "home-language" behind. Here the question is: What can a linguist do? I would prefer to quote Mosley [2007: ix] again: "Linguists who are outsiders must be sufficiently well trained as anthropologists, sufficiently observant and methodical as scientists, and sufficiently compassionate and sensitive as human beings, to be able to tackle both of these problems head-on. When a language is on the threshold of extinction, its speakers may well be demoralised in other, non-linguistic ways as well – economically deprived, dependent on aid, malnourished, unable or unwilling to draw on their cultural or religious traditions: any combination of these factors is possible. While it is not reasonable to expect the linguist to provide for all of these needs, it is impossible to act as if one were unaware of them."

Ostler [2005: 527-8] has made an interesting observation: "In the world's top twenty, all the languages have their origin in the south or east of Asia, or in Europe. ... What does account for their growth, then? It is noticeable that a great many of the languages (nine out of twenty) are spoken in the civilizations sustained by rice as a staple crop (Bengali, Japanese, Korean, Wu and Yue Chinese, Javanese, Tamil, Marathi, Vietnamese). Evidently, rice is capable of supporting dense and extensive populations, and its cultivation, through controlled flooding, requires a high level of organisation. Other languages which are not predominantly in the rice area are spoken in the neighbouring areas that have assumed political control of the rice areas (Mandarin Chinese, and Hindi and Urdu, which are linguistically in direct continuum if they are distinct at all)". If we add Telugu, which belongs to the rice-eating area and has been left out by Ostler, to the above list, 50% of the world's top twenty languages are spoken in the area where rice is a staple food. These interesting observations strengthen the above hypothesis that the relationship of "rootedness" to a land is a key factor in preservation of languages. The implication is that a change in the physical environment will obviously alter the linguistic and cultural environment of the concerned society. Due to large scale migrations from villages to towns and cities as well as rapid spread of the urban pop culture, the second generation everywhere has neither a "home-land" nor a "home-language". They are a part of the "motley mainstream" which is essentially similar all over the world.

I should mention here that the older generation people, especially the old women, living in the remote Indian villages are not at all familiar with the English calendar. Their life moves according to the traditional Indian calendar which most of the modern citizens do not understand. Many of the latter

people also find it difficult to name more than five flowers, five fruits, five birds and five animals in their own languages. They may know more words for these in English or Hindi; but they might have not seen most of them. Here I should cite what Whorf [1956: 135] observed regarding the cause of fire-accidents in the U.S. those days: "My analysis was directed toward purely physical conditions, such as defective wiring, presence or lack of air spaces between metal flues and woodwork, etc., and the results were presented in these terms. Indeed it was undertaken with no thought that any other significances would or could be revealed. But in due course it became evident that not only a physical situation *qua* physics, but the meaning of that situation to people, in the start of the fire. And this factor of meaning was clearest when it was a LINGUISTIC MEANING, residing in the name or the linguistic description commonly applied to the situation. Thus, around a storage of what are called "gasoline drums", behavior will tend to a certain type, that is, great care will be exercised; while around a storage of what are called "empty gasoline drums," it will tend to be different – careless, with little repression of smoking or of tossing cigarette stubs about. Yet the "empty" drums are perhaps the more dangerous, since they contain explosive vapor. Physically the situation is hazardous, but the linguistic analysis according to a regular analogy must employ the word "empty", which inevitably suggests lack of hazard." What it implies is that one's language has a significant influence over one's thought. I can mention a fascinating news that was telecast in some Indian television channels a few years ago. In a Madhya Pradesh village, the animals called "Neel gaay" in the local language, literally meaning 'blue cow', were happily destroying the crops and no villager was even trying to prevent them from doing so simply because they had the word "gaay" meaning 'cow' in the name. In fact, it was not a cow at all. So the state government was planning to change the name of that animal in order to avoid the problem. Following this logic we can argue that these people for whom it is hard to name more than five flowers, five fruits, five birds and five animals in their own languages should perceive the natural environment around them as an undifferentiated and not-so-useful continuum. If this argument is accepted, then how can we expect these people to be sensitive to and concerned about the conservation of bio-diversity? For these reasons, a sound knowledge of the mother tongue is crucial for appreciating bio-diversity and protection of environment. Not only that, the native words for lower order numerals, kinship, and body parts have been replaced to a regrettable extent by those of English and Hindi in these people's language.

Needless to state that diglossia is a part of life for most of the indigenous linguistic communities in this country. They speak the L(ow)-language for affective and social functions and the H(igh)-language for prestige-oriented

and informative functions. When a community starts using the H-language in the affective and social domains, it is a clear sign of endangerment. Thus, using English or Hindi words in the home domain is clearly a case of language loss. If this can happen to scheduled and state languages, the condition of minor tribal and indigenous languages can easily be imagined.

4. What Should Be Done?

Linguists can sensitise the citizens about the gravity of the problem and find the best possible ways to revitalise endangered languages and empower their speakers in consultation with other scholars and the endangered language speakers themselves. This brings us to the question of language revitalisation and empowerment. Giles [1977] has proposed the concept called "ethnolinguistic vitality" to predict the maintenance of a language by its speakers. He has emphasised three components:

a. the speech community's attitude towards their language,

b. the size of the community and its distribution, and

c. the institutional support for the language concerned.

I am not aware of any large scale study to determine different Indian language speakers' attitudes towards their own languages. But all studies on endangerment of Indian languages, whether major or minor, adduce evidence in support of a weak and/or negative attitude towards these languages. Needless to say that proper documentation and rank-ordering of various Indian languages with reference to their speakers' attitudes are absolutely necessary so that the languages at the bottom of the list can be immediately attended to.

The size of a community is undoubtedly a determining factor in preserving a language. We should better rephrase it as "concentration of speakers or speech-community in one place". Though the governments of Maharashtra, Odisha, and Tamil Nadu have to provide the required facilities for teaching Telugu to Telugu speaking children in the respective states, it need not be our primary concern. We must first concentrate on the status of Telugu in Andhra Pradesh itself where Telugu is the dominant language. At another level, maintenance of the Naga languages by the migrant Naga people at Delhi or Hyderabad is a secondary problem. The main concern is whether these Naga languages are maintained in their home-land, i.e. Nagaland. If not, revitalisation of such cases is a priority.

The third component of institutional support for language education in India is at best a parody of what it should have been. Two of my students, who had completed their school and college education in Odisha including studying

Odia as a subject in school, are not able to speak Odia. It is well known that tribal language education has never been a serious business in this country. Many learned people as well as teachers think that tribal languages are "dialects" because they are not written and that is why, they need not be taught or used as media in education. All these warrant a complete revamp of mother tongue/first language teaching at every level, and a thorough planning for tribal language education in this country. Though we hear a lot about status and corpus planning, what we need the most is "acquisition planning" (see [Kaplan and Beldant 1997] for a discussion on it) so that many language retention issues can be resolved and the gap between the home- and school-language can be bridged. Keeping these in view, a Mother Tongue based Multilingual Education Programme has been launched in India especially for the minor and indigenous language speakers. This Programme emphasizes the role of the mother tongue in education while other important languages are also taught to the students. Though it is confined to a few states only, it has shown very positive encouraging results.

Under the impact of globalisation, a number of changes are taking place in this country over which we have no control. Therefore, breaking down of joint families cannot be reversed. Movement of linguistic groups from their home-land to other language speaking areas for greener pastures cannot be checked. Interlingual marriages cannot be prevented. But we can do a number of things that will ultimately lead the endangered language speaking communities to empowerment.

The first and the easiest step is to record the data on each and every language spoken in this country in the decennial Census reports. The next significant step would be to provide these languages with scripts in consultation with the concerned communities, because it will give them an identity along with a proud feeling that their language is comparable to the so-called developed languages. We know that tribal communities possess a very rich oral literature and a vast wealth of alternative knowledge systems. Once a writing system is in place for each of these languages, the speakers may be encouraged to document these treasures, and in the process it will bring recognition and visibility to many in the society. Then, newspapers, books and magazines can also be published for the people speaking these languages which will make them involved in the activities of their own languages.

Radio and television are not only an integral part of every household today; they are the most powerful media to make or mar anything. With the kind of advanced technology available in India, it is quite feasible to start community radio and television channels for most of these indigenous linguistic groups. This will not only revitalise the domain of creative literature but also open new

entertainment avenues in these languages which, in turn, will act as a filter and prevent the speakers from getting softened towards other dominant languages.

Of late, the Government of India has initiated a project called the Scheme for Protection and Preservation of Endangered Languages (SPPEL) through the Central Institute of Indian Languages, Mysore. A first list of 520 languages spoken by less than 10,000 speakers has been prepared and documentation work on some of these languages has started.

The Language Division, Ministry of Home Affairs, Government of India has also started the Mother Tongue Survey Project under which a number of small languages are being documented and studied.

5. Concluding Remarks

To conclude, I would like to emphasise that every language, be it Toda in the Nilgiri Hills or Savara in the dense forests of southern Odisha, has a right to live its full life and is equal to any "international" language in this world. That is why Sapir [1979: 219], one of the founders of modern linguistics, had long ago remarked: "When it comes to linguistic forms, Plato walks with the Macedonian swineherd, Confucius with the head-hunting savage of Assam." I should close this discourse with what David Crystal [2000: IX-X] stated in his now classic book entitled *Language Death*: "All I know is that the issue (of language death) is now so challenging in its unprecedented enormity that we need all hands – scholars, journalists, politicians, fund-raisers, artists, actors, directors... if public consciousness, let alone conscience, is to be raised sufficiently to enable something fruitful to be done. It is already too late for hundreds of languages. For the rest, the time is now".

References

1. Baker, C. (1993). *Foundations of Bilingual Education and Bilingualism*. Clevedon: Multilingual Matters.

2. Baker, C. (2001). *Foundations of Bilingual Education and Bilingualism*. Clevedon: Multilingual Matters. (3rd edition)

3. Brenzinger, M. (ed.) (1992). *Language Death: Factual and Theoretical Explorations with Special Reference to East Africa*. Berlin, New York: Mouton de Gruyter.

4. Brenzinger, M. and de Graaf, T. (n.d.). Documenting endangered languages and language maintenance. Contribution to the UNESCO Encyclopedia of Life Supporting Systems (EOLSS) 6.20B.10.3.

5. Crawford, J. (1999). *Bilingual Education: History, Politics, Theory and Practice*. Los Angeles: Bilingual Educational Series. (4th ed.)

6. Crystal, D. (2000). *Language Death*. Cambridge: Cambridge University Press.

7. Evans, N. (2010). *Dying Words: Endangered Languages and What They Have to Tell Us*. MA and Oxford: Wiley-Blackwell.

8. Finegan, E. and Besnier, N. (1989). *Language: Its Structure and Use*. San Diego: Harcourt Brace Jovanovich.

9. Fishman, J. A. (1978). Positive bilingualism: some overlooked rationales and forefathers. In: *International Dimensions of Bilingual Education*, ed. by J Alatis. Washington, DC: Georgetown University Press, pp. 42-52.

10. Fishman, J. A. (1991). *Reversing Language Shift*. Clevedon: Multilingual Matters.

11. Garcia, O. (1992). For it is in giving that we receive: A history of language policy in the United States. Paper presented at the conference "American Pluralism: Toward a History of Discussion" held at SUNY, Stony Brook on 7 June 1992.

12. Giles, H. (ed.) (1977). *Language, Ethnicity and Intergroup Relations*. London, New York: Academic Press.

13. Grenoble, L. A. and Whaley, J. (2006). *Saving Languages: An Introduction to Language Revitalization*. Cambridge: Cambridge University Press.

14. Gumperz, John J. and Wilson, R. (1971). Convergence and creolisation: a case from the Indo-Aryan/Dravidian border in India. In: *Pidginization and Creolization of Languages*, ed. by Dell Hymes. Cambridge: Cambridge University Press, pp. 151-167.

15. Harrison, K. D. (2007). *When Languages Die: The Extension of the World's Languages and Erosion of Human Knowledge*. Oxford: Oxford University Press.

16. Holmes, J. S. (2000). The name and nature of Translation Studies. In: *The Translation Studies Reader*, ed. by L. Venuti. London and New York: Routledge, pp. 172-185.

17. Kluyev, B. I. (1981). *INDIA National and Language Problem*. New Delhi: Sterling Publishers.

18. Krauss, M. (1992). The world's languages in crisis. *Language* 68, pp. 6-10.

19. Lo Bianco, J. and Rhydwen, M. (2001). Is the extinction of Australia's indigenous languages inevitable? In: *Can Threatened Languages Be Saved?,* ed. by J. A. Fishman. Clevedon: Multilingual Matters, pp. 391-422.

20. Mackey, W. F. (2003). Forecasting the Fate of Languages. In: *Languages in a Globalized World*, edited by G. Maurais and M. A. Morris. Cambridge: Cambridge University Press, pp. 64-81.

21. Mohanty, P. (2008). The other maternal uncles in Indian languages. In: *Ethnographic Discourse of the Other: Conceptual and Methodological Issues*, ed. by Panchanan Mohanty, Ramesh C. Malik and Eswarappa Kasi. New Castle: Cambridge Scholars Publishing, pp. 69-90.

22. Mosley, C. (ed.) (2007). *Encyclopedia of the World's Endangered Languages*. London and New York: Routledge.

23. Myers-Scotton, C. (1992). Codeswitching as a mechanism of deep borrowing, language shift, and language death. In: Brenzinger (ed.), 1992, pp. 31-58.

24. Ostler, N. (2006). *Empires of the Word: A Language History of the World*. New York: Harper Perennial. (first published in 2005).

25. Pattanayak, D.P. (1981). *Multilingualism and Mother Tongue Education*. Delhi: Oxford University Press.

26. Sapir, E. (1979). *Language: An Introduction to the Study of Speech*. London: Granada. (first published in 1921).

27. Schiffman, H. F. (1996). *Linguistic Culture and Language Planning*. London and New York: Routledge.

28. Schmid, M. S. (2002). A new blueprint for language attrition research. In: *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues,* ed. by M. S. Schmid, B. Köpke, M. Kaijzer and L. Weilmer. Amsterdam, Philadelphia: John Benjamins Publishing Co., pp. 349–363.

29. Whorf, B. L. (1956). *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf* (ed. by J. B. Carroll). Cambridge, Mass.: The MIT Press.

Claudia WANDERLEY

Professor, Researcher,
Centre for Logic, Epistemology and the History of Science,
UNICAMP University
(Campinas, Brazil)

To Map Initiatives/Research on Multilingualism in Brazil: An Approach to Preserving Cultural and Linguistic Identity

In spite of their occasional interest in touching the other of the West, of metaphysics, of capitalism, their repeated question is obsessively self-centered: if we are not what official history and philosophy say we are, who then are we (not), how are we (not)?

Gayatri Spivak, 2006

For historical reasons the co-existence of many languages in one politically unified territory is not a part of linguistic studies in the majority of countries of the world. Multilingualism [and therefore multiculturalism] is a new subject in post-colonial context. The studies of language description and differentiation are well developed in most countries. But studies about linguistic plurality, multilingualism, and the development of social analysis concerning international policies, and the perspective of a common research agenda among researchers haven't formally started. This work proposes the reverse question of European academic historical epistemological path; it aims 1) to comprehend linguistic plurality and digital inclusion of local languages in Brazil, 2) to map the researchers working on the subject and the fields of knowledge that they represent. We present some of the results of the first year of activities of this project, which has the support of São Paulo Research Foundation (FAPESP). For the next year local projects shall be indexed, visible and available for researchers interested in Multilingualism and Multiculturalism through an observatory. The main result shall be a first compendium of the body of knowledge that involves the available initiatives on Multilingualism in Brazil, and a first analysis of the presence of this multidisciplinary theme in transnational academic tradition, concerning Brazilian data.

1. Presentation

Multilingualism might be differently understood whether in a western paradigm – meaning European understanding of multilingualism – or if we are in an interpretative key that considers ethnic and cultural erasure and/or

exploitation by a dominant. In the case of this paper, concerning a first mapping of multilingualism in Brazil, there is a preamble which is worth mentioning, only to assure that the notion of multilingualism won't be taken for granted (or to be read as analogous to the European approach to multilingualism) in our context, our historical, social, economic, and – most of all – educational situation in Brazil. As this text is thought for a broader audience, I understand that this preamble should bring forward a bit of information about our historical relationship with language and culture since the "appearance of Brazil (of what would become Brazil) in Western narrative. We are writing in a post-colonial critic perspective, having as main references authors like Gayatri Spivak, Homi Bhabha, Said Ali and Franz Fanon.

I present here a minimal sketch of a timeline only to introduce a bit of Brazilian historical context concerning multilingualism, in the way that I understand it today. In 1452 the Portuguese Crown authorized to reduce the conditions of Africans to slavery aiming their christianization, with the support of the institutionalized Church, the Pope, registered in the bills *Nicolau V Dum diversos* and *Divino Amore Communiti*. In 1492 Christopher Columbus' fleets appear in American continent, after some storms, "thinking" that they had arrived in India. The first document that alluded to our existence is the Treaty of Tordesillas in 1494, celebrated among Spain and Portugal who divided the "discovered" lands among them. But the name of the continent "America" is inherited from Americo Vespucio, who arrives on our shores in 1497. The "discovery of Brazil" is dated in 1500, and credited to a Portuguese called Pedro Alvares Cabral. Until the year 2003 the only version of the origin of the country in school books would be the narrative of arrival of the Europeans on our shores, with total disregard to the approximately 1300 indigenous different ethnic groups [Aryon Rodrigues 2002] that lived in the region that later would be called Brazil.

Of course the expeditions of European kingdoms to our lands through navigation were not envisaging cultural or linguistic exchange. It was, and still is about exploitation and profit. The real discovery of Portugal[104] is that the business of slavery was much more profitable than the commerce of other goods. And when the king realizes it in 1605, the Portuguese Empire shall combat any exchange of slaves, or goods, within their territory – especially on the "colonized" ground – not related to the metropolis (Lisbon), or in other words all the commerce in the Portuguese empire has to pay taxes to Lisbon. In this sense the exploitation and colonization processes are coincident to the market lessons which will build the market of slavery established by the Portuguese King. The introduction of Africans in America allied to the

---

[104] *Cf. Alencastro 2000.*

embargo of the captivity of the Indians, forced the local farmers to export their goods to the metropolis as well as to reach for the metropolis to import their *factors of production* (slaves).

The exchange among the kingdom and the colonies is balanced with the slave trade, and Portugal becomes a Global Trader. According to Luiz Felipe de Alencastro, "[...]the slave trade is not reduced to the commerce of black people. It brings decisive consequences to the formation of Brazilian History, the trade extrapolates the register of the operations of buying, transportation and sale of the Africans to mold the economy, the demography, the society, and the politics of the Portuguese America."[105] [Alencastro 2000: 29]. The Atlantic trade system might be shortly described as the robbery of African peoples and the slave based agriculture in America. The commerce of slaves is the biggest source for the king's treasure, and the taxes of the trade surpass the economic gains of slavery.

There are many levels in which the structure of deportation of Africans is determined by the capitalism of commerce, and Alencastro considers the following levels: 1) the metropolis controls the trade, and therefore has the command of the slavery system; 2) the Crown and its administrators find new funding for the kingdom creating several taxes for the trade; 3) the arrival of the Africans alleviates the indigenous captivity, what shortens the autonomy of locals over indigenous slavery, and alleviates the tensions among the jesuits, locals and the king administration; 4) the traders will combine the advantages of the trade of sugar with the advantages of the trade of slaves; 5) the commerce abroad is enhanced, the profits from the farms are in part to buy more slaves to expand. The slaves represent 1/5 of the investment in a cane sugar mill, and half of the investment of the sugar cane farmworkers, says Alencastro. And 6) on the long run, buying Africans is favorable for the locals, because the survivors of the trade process are the stronger and have passed for an intense "desocialization". This last level is exactly on the reverse movement of our intentions, for this intense desocialization mentioned by Alencastro is exactly the loss of roots, and ethnic and community values. Language and culture are luxuries unattainable, concerning the degree of dehumanization promoted to non-Europeans.

It is unnecessary to underline that the decision to de-humanize Africans and Native Americans, to transform human beings of the African continent into merchandise, to strongly invest in the monopoly of the commerce of human

---

[105] *"[...]o trato negreiro não se reduz ao comércio de negros. De consequências decisivas, na formação histórica brasileira, o tráfico extrapola o registro das operações de compra, transporte e venda de africanos para moldar o conjunto da economia, da demografia, da sociedade e da política da América portuguesa." [Alencastro 2000: 29]*

beings (Africans) in a business society between the Portuguese empire with the catholic church, and to admit slavery process and slave commerce as the most lucrative business in the colonial process are some of the symbolic heritage that we first have to deal with – in our post-colonial context – if we are serious about considering linguistic and cultural identity.

In a first attempt for understanding the process of preserving cultural and linguistic identity of local languages and local cultures in Brazil, the main notions that come to mind are *erasure* and *desocialization*. A consistent erasure promoted for centuries, legitimated by history, religion and academic knowledge, and some centuries of existence of a narrative about our communities in Western knowledge, without a real concern to our reality. In our hypothesis, this erasure is possible due to a process of naturalization of the dehumanization (desocialization) of native peoples and African peoples promoted by the Europeans, to enable exploitation and profit. This is our starting point to consider multilingualism and multiculturalism in Brazil.

2. Multilingualism: Questions

Multilingualism understood as a transversal subject is quite new in academic environment. In fact, in Portuguese speaking countries, it is still an inexistent concept, which has to be created in partnership with the most needed communities that have no idea of the effect of such an abstract object in their lives. A project on Multilingualism in Portuguese speaking countries, as this one, is settled on historical paradoxes, or at least interesting questions. How to promote multilingualism in public universities through a language that has been historically the monopoly of the local elite? How could we reach these local communities within our countries while a great part of the efforts in national academies, and the induction of national funds for research, is for us to be abroad publishing with researcher teams of the USA or Europe?

The objective of becoming part of what we call here the first world of academic production is in strange relation to the contact with our local reality, especially in humanities. To understand the the working class, for instance, we need to make an effort to zoom out from the illegal third sector, that is huge in countries in development as ours. Or thinking about education, and schools, it is necessary to put an effort of re-projection, in our public institutions, of the value printed in Europe or the USA.

To keep our disciplines in order with international debate, we must reaffirm our bonds to what authors in the Northern hemisphere thought about us. To inherit gratefully the academic debate from this point. Is it possible? On

multilingualism issues, mostly in post-colonial countries? We believe so, but in a particular way, mainly enhancing open debate in the Southern hemisphere, sharing open texts, enhancing free access to digital academic content, and broader circulation of local knowledge.

1) The Portuguese Language

One strong discursive focus on Portuguese colonization has been on the relationship between a unified language and the unity of the territory of the Portuguese empire. Since it was not possible to visualize the extent on geographic level, language as a powerful asset of national state was elected as the bond, enabling conquerors to unify and [de]territorialize these many different lands. The Portuguese language becomes a language that would open doors to all continents. This language, for the dominant, promotes a unified territory and at the same time is considered to be a door to many continents. This paradoxical characteristic, a unified territory that is also many different ones, becomes part of our linguistic legacy.

Nowadays, after the independence of seven former colonies, the Portuguese language became the official language of seven national democratic republics, seven Nations. And following its historical pattern of producing its territory and its paradoxical doors for his official owners, it pointed to the national (and post-colonial) unities. The doors now are among nations in all continents, and curiously a new door appeared: the Portuguese language now also leads to Europe. If we keep standing on this metaphor, little pathways appeared within territories, usually incomplete pathways. The language of the new State did similar job with local communities, or tried to. It should arrive at their territory and unify it, which is not always a successful agenda. Some of the nations recognized other national languages, but the official language, the language of the State, is still Portuguese.

In Brazil, the national language and the official language are called Portuguese, and we have around 200 original languages in the country which are not recognized. But in other Portuguese post-colonies this difference is at stake, there are local languages on the street and there is Portuguese language only at official level. In parallel and to get closer to our "brothers and sisters" in the post-colonial situation, it is possible as well to consider Vernacular Brazilian Portuguese (the official level) as distant from the Brazilian Portuguese of the streets as a second language would be. Eminent Brazilian sociolinguists, as Prof. Stella Maris Bortonni (at UnB university), propose that Vernacular Portuguese should be taught in schools with the

methodology used for second language learning. Once we admit that, our gaps and/or perceptions of educational environment have a lot in common with many post-colonial situations.

Nevertheless for those communities that have as a mother tongue a non-official or non-national language there is a strong social problem. The questions that we bring forward are related to the educational reality of these social groups, considering their representativity and visibility and technical and cultural resources available as part of our society. Of course, and preferably, all social groups are very different from each other, and the aspects in multilingualism and multiculturalism are right at the frontier of unity and diversity. In this perspective public policies on linguistics and on telecommunications play a major role in our/their possibilities of expression and information access.

As we understand here, the strongest role of the Portuguese language is to guarantee a territorial asset for the State, and its prior direction is unity. Linguistics in this case is to be understood in close relationship, possibly in a metaphoric substitution, to geopolitics. And the early debate on multilingualism and multiculturalism in our grounds is a small feather in the balance of truth of so many heavy hearted centuries.

2) Multilingualism in Portuguese Speaking Countries

> *Le rêve de ceux qui rêvent concerne ceux qui ne rêvent pas, et pourquoi cela ça les concerne? Parce que dès qu'il y a rêve de l'autre, il y a danger. A savoir que le rêve des gens est toujours un rêve dévorant qui risque de nous engloutir.*

> Giles Deleuze, 1987

Although in this paper multilingualism and multiculturalism are comprehended as resources for our nations and for a trans-national network, multilingualism in such historical environment, and particularly in Brazil, is comprehended as a flaw in territorial unity, and an inability of the government to promote civil education and nation-state values.

The Knowledge Society, that would come from the Information Society, is a dream we have learned about at UNESCO headquarters. Isn't it wonderful to beat illiteracy, bad educational systems and lack of libraries or educational resources, all at once with digital world? More than that, in our present situation, for Portuguese speaking countries, it is a moment to regain contact

with each other, to comprehend our history, our academic world is flourishing in a new collaborative perspective.

But there's a tension in this group, for Multilingualism was never a subject amongst us – the academics of Portuguese speaking countries. *In fact I'd say that multilingualism is the impossible subject of all geopolitics of Portuguese colonization. And I believe it should be one of the starting points of postcolonial criticism concerning our countries.* So its presence is marked with silence. And that's what happens when we gather in our meetings, and the subject is multilingualism: silence, a discourse of the negative presence of this object. And many reactions came from this. In Brazil some colleagues felt it as a hopeless situation, and tried desperately to fill this linguistic «gap» with their previous works on different subjects, [obviously] without success.

But as a linguist, as a discourse analyst, I understand that it is a situation that puts us all in a very meaningful space that lies before the sign, before language practice. This same space could be interpreted as a place of annulment. It is a political choice to choose the potential of this silence that bears innumerable possibilities for research. I comprehend this tension as a good symptom, meaning that there's a lot of work to do together concerning the original proposal of the UNESCO Chair on Multilingualism in Digital World created at Unicamp University in 2007. Still the academic ideological theater is filled with silence. This is our statement at this moment. There is a group, we are interested in the subject, we are writing and working on individual and common projects, we promote international meetings, and we still bear our silences. Meanwhile we have produced two books.

But although it is not a traditional academic subject, multilingualism has been discussed in Brazil, ever, outside academy. Indigenous peoples are affected by multilingualism on their everyday lives, putting it in cause very frequently. These discourses haven't reached the academy yet. But recently [ten to fifteen years approximately] in Brazil, there have been Higher Education programmes specially shaped for indigenous societies, and an effort to have graduated teachers in their own villages permitting criticism in educational practice. But it is a long way to find these initiatives inside the country and to ask about their interest to promote a collaborative debate.

The reason for which I decided to map the researchers working on the subject, and the fields of knowledge that they represent, is because in 2007, when I brought to Unicamp University the proposal of the UNESCO Chair Multilingualism in Digital World as a result of my PostDoc in France, there were no immediate partners/peers to structure the debate on multilingualism in digital world in Brazil. And it was not evident to which path should we

turn to, as a research group, since we became a research group with a strong UNESCO role[106], through *Humaniredes,* an initiative of Prof. Frances Albernaz (at that time working as a UNESCO Officer in the field of Culture at UNESCO headquarters), and with the support of the UNESCO National Commissions from Brazil, Cape Verde, Saint Tome and Principe, Bissau Guinea, Angola, Mozambique, China, East Timor, and Portugal. So, with a group of fourteen higher education institutions and a common heritage of post-colonial linguistic issues, along with Prof. Claudio Menezes (at that time working as a UNESCO Officer responsible for Multilingualism in Digital World at UNESCO headquarters) we decided that although the proposal was in a network structure, it would be wiser to submit the proposal of a UNESCO Chair, and not a UNESCO UNITWIN (the idea – lately we understood that was a wrong assumption – was that a UNESCO Chair would be easier to administrate than a UNITWIN). It is important to stress that in this network we included the former colonizer and the post-colonial countries; what brought us a lot of experience with post-colonial issues, and made us understand the difficulty to deal with this team without a post-colonial criticism framework established to better comprehend the colonization procedures entangled by the Portuguese in our countries, the power switch promoted by our independence as nations (in different periods) and its effects. We had quite a turbulent experience with the first Chair at the University of Campinas in Brazil, in national and international grounds. But we gained a glimpse of a huge agenda, and we were able to foresee what kind of research is ahead of us if we really want to deal with this theme. So the reflexion, presented here, is a result and a demand of this former experience, as well as another intent to better understand the Brazilian research profile related to multilingualism and multiculturalism. Hopefully we will be able to better understand our real possibilities to expand, and best ways to interact with researchers in Brazil.

Prof. Daniel Pimienta, researcher from FUNREDES, has stressed the need to map not only researchers in academic environment, but in civil society initiatives as well. I hope that for the next IFAP meeting on Multilingualism in Digital World we can show him some good news. But for now we will stick to Brazilian national academic environment.

---

[106] *It is important to express our gratitude to Frances Albernaz, UNESCO officer, who managed to gather multilingualism and multiculturalism issues in Portuguese Speaking countries through her network established in UNESCO called Humaniredes. The goals of Humaniredes are a strong inspiration and in fact they created the bonds for our research network, and Frances' ideas shaped and still have a major role in the research that I develop today on multilingualism.*

Of course, to develop a research on multilingualism in Brazil we shall deal with a very specific context for multilingual and multicultural issues. First of all, Brazil – as well as Portuguese speaking countries – has historically strong monolingual policies (and consequently a monolingual imaginary in academy). This is the first situation that a researcher who wants to work with multilingualism has to deal with: the subject is not evident, or better, the subject is invisible for most part of the researchers or funding institutions that would normally be able to support or interact with your research. It is common sense – of course a common sense historically produced – that we (Brazilians) only speak Portuguese, and lately we recognize two languages for the state: Portuguese (recognized as a state language in 1758) and Brazilian Sign Language (recognized as a state language in 2002). No other language is recognized as a national language in Brazil, although we have around 200 living languages from original peoples (all of them at the brink of disappearance). So, with such a rich linguistic environment, monolingualism is, and always was, the main dish served in K-12 schools and in national, public and private universities.

Why a Brazilian linguist would want to work with these notions [multilingualism and multiculturalism] in such national historical adverse circumstances for the theme? In my own experience it looked like a relatively new academic approach for the subject in Brazil. And specifically from this perspective it was a nice academic move, to start fresh in the comprehension of multilingualism in digital world. But it was not so simple as I thought, for Brazilian academy is extremely traditional and any novelty is easily understood as lack of respect for the former disciplinary area, with strong and violent reactions at the institutional level, even reaching personal level as well. This kind of rejection of new debates [and we could easily add the rejection of new bibliographies] is also related [in Brazil] to the militarization of public institutions, public universities included. The military dictatorship (1964–1985) has left us – among other things – the heritage of authoritarianism and an awkward notion of safety based on the exclusion of anything that does not agree with the actual authority. Such attitudes are currently part of Brazilian institutions, and unfortunately they play a strong role in our actual academic life. So, here is a second situation to deal with; the study of multilingualism in digital world in Brazil will certainly promote a new configuration of the knowledge related to the subject, and the effects of such reconfiguration are not obvious.

3. Multilingualism and Multiculturalism in Digital World

Digital world is not a discipline, neither is multilingualism or multiculturalism. But when we bring the debate to the academy they "naturally" are expected to become part of a knowledge field, and they are often [mis]interpreted as

such. At this moment calling them transversal doesn't say much about our immediate interest, or about our actual needs, although it says a lot about the kind of partners that we are interested in – just everyone who feels related to it.

We have proposed to the Brazilian institutions dedicated to fund national and regional research to consider multilingualism in communication theory, and FAPESP (the funding agency responsible for this project) considered it in this area. This goes along with the place of this debate in UNESCO, which is the Communication and Information Sector, and it also permits us to bring essential issues that are on the brink of this proposal: liberty of expression, democratic access to information, social inclusion, digital inclusion; availability of editor houses for minority languages, translation of important literature to local languages, translation of important local discourses to broader access, and after all: how to have direct access, for instance, to Cape Verde or/and Indonesia academic production? How could we be aware of our colleagues' works in progress, to be able to collaborate? What is the research going on in Madagascar? What are the Mozambican researchers debating about now? Our first necessary activity is to figure out how it is possible to develop communication and information among us, out of the American and European broadband highway imaginary. In fact, if we are proposing a South-South collaboration, of course we don't always have broadband, or computers, or electric energy, but we always have knowledge production. It is obvious that our cultures and our public research institutions are at work, there's no such people that doesn't produce and make their knowledge circulate with great effectiveness. So, the question is how to recognize, make publicly available and exchange our knowledge production?

Unfortunately, or fortunately, to comprehend this subject we are touched by Marxist or Althusserian logo-centrism as a pattern of domination. And this can be a blindfold to other possibilities of communication and knowledge circulation. This is not a reference to non verbal corpora only, it is mainly a trace to be considered: because we are so implicated in our texts and looking for scientific publications, and there might be other possibilities of knowledge circulation that are not in our immediate logo-centric and digital-networking sight. Also this curiosity shouldn't be interpreted as a denial of current initiatives, and actual immense effort to textualize and digitalize knowledge. On the contrary, it is just an attempt to embrace new options, to recognize our traditional and local ways of communication and information spread, and to be able to maintain a conversation somehow within and beyond, or besides, *textology*.

The economics of a digital text is an interesting subject in this perspective, because it is determinant to our working conditions in digital world. Mainly

because it is very expensive to put the script of a language in Unicode, then to translate platforms and software to this new language, to provide manuals, technical assistance, and finally to find an interesting number of consumers to respond to the need of this digital linguistic infrastructure. Nowadays we have around six thousand languages in the world, the African continent solely bearing two thousand. On the Internet, we have around sixteen languages functioning in their whole technological capacity, and around sixty languages being able to resist well on new ICT software. The others which are available are usually lending the infrastructure of another linguistic system.

According to the Ethnologue [2014] 96% of the languages of the world are spoken by 4% of the world population. It seems that the economy and possible public policies won't fit for the majority of language speakers, when we consider ICT industry and its real possibilities. This being said, we have two options: to develop bilingualism to access digital platforms and therefore digital data, or to investigate the possibility of digital smoke signals to start a digital conversation; both are very good options. Basically it means, when choosing bilingualism, to look for free software and open resources and the collaborative spirit, and when we choose "smoke signals" this possibility depends basically on a collaborative, creative and proactive network. It means on short terms to consider using well known digital objects in different communication functions to include local languages and preserve local cultures. We are working with both in different scales.

Another interesting point that I'd like to bring to the understanding of "multilingualism in digital world" is that it needs a multidisciplinary approach (in short, multilingualism is not always for linguists). There is a large spectrum that goes from language acquisition, computer programming, bilingual/multilingual education, free and open software, public policies on ICTs, language revitalization, ethno-disciplines, regional and national recognition of linguistic and cultural heritage, anthropology, data modeling, arts, etc. There is a myriad of areas that could interact with the subject, because at the end we are talking about human values and practices that we don't want to disappear, and we want to enhance multilingualism and multiculturalism in any way that might benefit local languages and local peoples life; and the digital world is full of possibilities to achieve that. In this sense, there are always ethical aspects to be considered.

1) Some Brazilian Data about Multilingualism in Brazilian Academy

Brazil had its first scientific and cultural institutions of higher education in the 19th century, with the arrival of the Portuguese royal family who

fled from Napoleon invasion. The arrival of the royal family promoted the creation of educational institutions. But the first Brazilian University dates from 1909 at Manaus in the state of Amazonia. So the institution university is relatively recent in Brazil, always strongly related to the promotion of elite interests. My generation of researchers [we are in 2014, I had my PhD in 2003] is the first generation of Brazilians that have a public sponsorship from the Brazilian state to study through higher education (Master and PhD) without family economic support. So, we are a first generation of Brazilians in academy that do not owe to their families or to a specific social group the investment in our education. We had public funding to support our education. It practically means that we are allowed [or at least we are not constrained] to pose questions.

So, to start to understand who we are in terms of a real possibility of debating multilingualism, I looked at Lattes platform (lattes.cnpq.br/), at The National Council for Scientific and Technological Development (CNPq), which is an agency linked to the Ministry of Science and Technology (MCT). All the researchers working on national ground are encouraged to upload a Curriculum Vitae in Lattes. So it gives us an idea of who is actually working in Brazilian academy. If we get into the platform and ask for PhDs data, it will show that today there are 109,799 PhDs in this country with a total population of 200 million people. If we look for a specific area concerning language we shall find 2,000 PhDs in Linguistics, "Letters" and Arts. The group of PhDs directly related to languages is as big as a full conference house, and extremely disproportional to the size of the country, the size of our population, and the actual need for research in our area. Also this data explains quite well the possibilities for forming groups – unfortunately very common in our reality – with a familiar working profile, and not a public knowledge production profile, meaning endogenous groups, or feudal group structures, or even as we created an expression for it – at the backstage – authors' "little churches" (in Portuguese we say "igrejinhas"). This "professional" profile, bringing family bonds in their professional bonds, is usually against any novelty. Sérgio Miceli in a sociological study traced the profile of older Brazilian intellectuals and mentioned a profile of double dependency [Miceli 2001: 59], they are dominated by inner relations of forces of the oligarchic group, for they don't really belong to this group; and they also belong to a dominated field, in the position that they occupy in international intellectual relations.
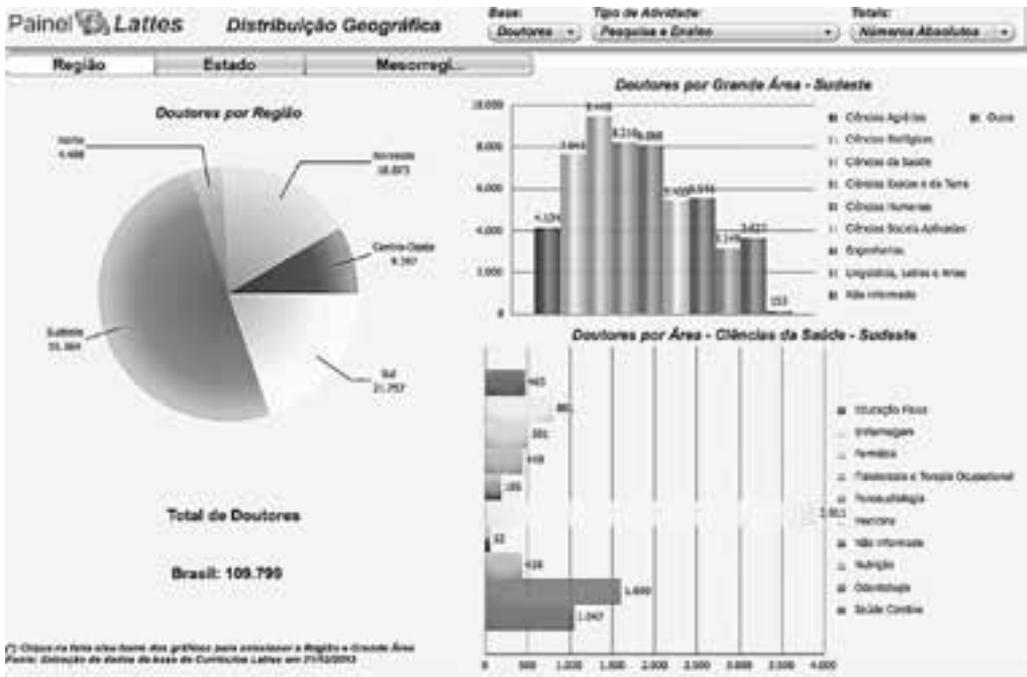
Another public platform that is worth consulting is the Digital Library of Brazil of Thesis and Dissertations (BDTD) (http://bdtd.ibict.br/), hosted at IBICT – The Brazilian Institute on Science and Technology (Instituto Brasileiro de Informação em Ciência e Tecnologia), an agency linked to the Ministry of Science and Technology (MCT). Since 2010 in Brazil it is mandatory for public institutions to upload at BDTD all the academic degree production derived from public funding. The library and the full texts are available for free access. So we have today the following numbers: 179,038 for Master Degree or Dissertations + 65,869 for PhD or Thesis, meaning the total available production of 244,907 documents responsible for academic degrees. In the table below we can see the number of works uploaded every year.

If we search on both platforms for the keyword multilingualism (multilinguismo) we will find: Multilinguismo/multilingualism

Lattes: 288 researchers, 166 PhDs within this ensemble

BDTD: 28 documents

And if we look for the keyword multiculturalism (multiculturalismo) we will find:

Multiculturalismo/multiculturalism

Lattis: 4,568 researchers, 2,165 PhDs within this ensemble

BDTD: 167 documents
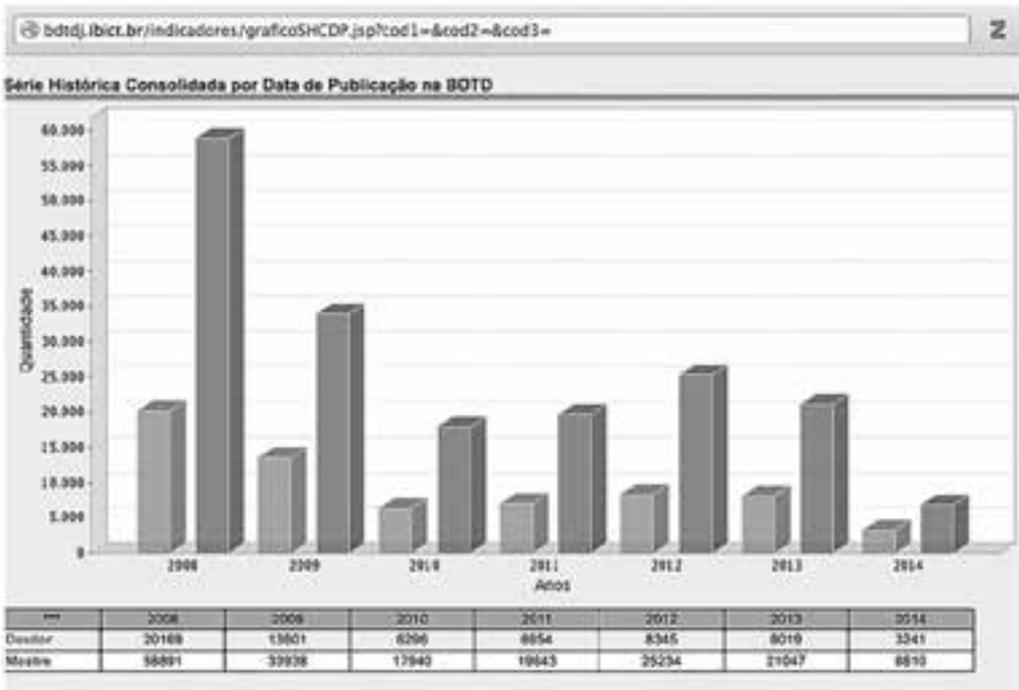
We might make some considerations of why a country with such a linguistic diversity would practically ignore the subject, if we understand what the choice of keywords to register in both platforms indicates. There is a possibility that our colleagues are unaware of the terms multilingualism or multiculturalism, and therefore their choice of keywords sent they apart from us. Another hypothesis is our funding agencies' effect. For we have in Brazil funding agencies for research that promote mainly a deduction process to support research, it means that they choose what will be funded and the researcher has to adapt if she/he wants to have some financing. More in the field of the imaginary, we have a historical monolingual policy, that might enhance a monolingual imaginary in academy and therefore in research perspectives, or we might as well consider that it is a relatively new academic approach for the subject in Brazil, and also that it is far from mainstream and the use of such keywords would not make much sense.

I have used the open software *lexico¹⁰⁷* (http://www.tal.univ-paris3.fr/lexico/) to start analyzing the data of lattes. When analyzing the results for 166 PhDs that consider multilingualism as one of the keywords of their work, we have the following disposition by field: 82 researchers in Linguistics (including 21 in Applied Linguistics), 44 in Letters (Letras), 14 in Education, 11 in Literature, 5 in Culture, 4 in Communication and Psychology, 3 in Anthropology, Philosophy and History, and 1 in Psychology and Interdisciplinary. And the degree of proximity among the researchers is not very high, as we can see in the next table of a factorial analysis of the data concerning each of the 166 individuals:
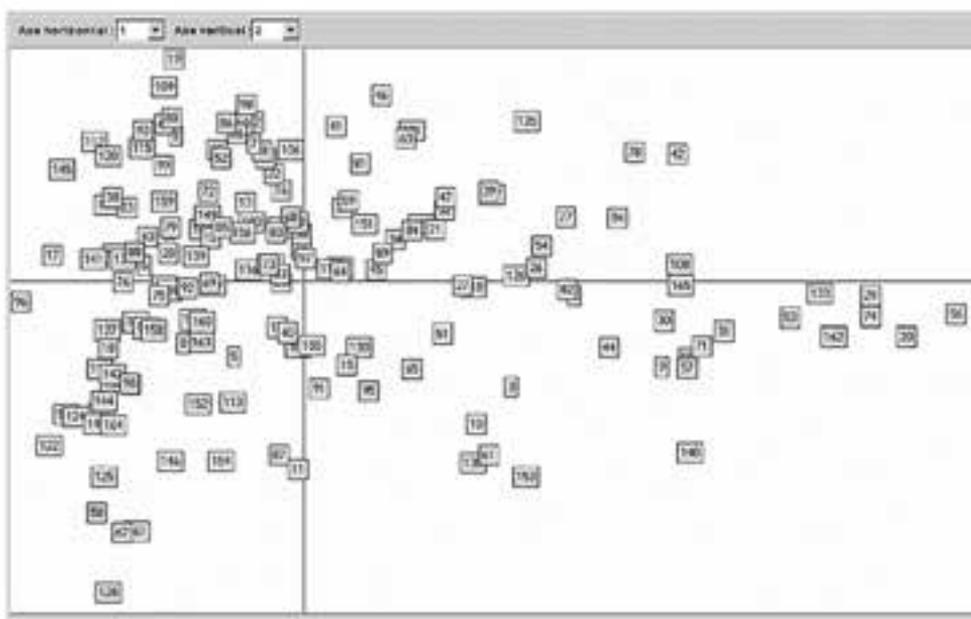
(each blue square represents a researcher)

So in terms of a first approach to this data we cannot affirm that we have an established (even though small) research field. And we cannot affirm that they

---

¹⁰⁷ *"Lexico3 is the 2001 edition of the Lexico software, first published in 1990. Functions present from the first version (segmentation, concordances, breakdown in graphic form, characteristic elements and factorial analyses of repeated forms and segments) were maintained and for the most part significantly improved. The Lexico series is unique in that it allows the user to maintain control over the entire lexicometric process from initial segmentation to the publication of final results. Beyond identification of graphic forms, the software allows for study of the identification of more complex units composed of form sequences: repeated segments, pairs of forms in co-occurrences, etc. which are less ambiguous than the graphic forms that make them up."* *(http://www.tal.univ-paris3.fr/lexico/).*

are mutual peers. The data considered for this first interpretation were: PhD academic degree, year, academic position, main area, keywords, institution, city, region, country, funding agency.

2) *The Internet as an Environment for Preserving Cultural Identity*

Possibly because Brazil has a common belief of being a melting pot of cultures, and this is something to be proud of, the idea of multiculturalism has spread a little bit more. In this sense I made an experience to work with multilingualism through the key of multiculturalism. And as I wished to keep on working with notions like information and technology, I have proposed an approach to the notion of culture based on systems of information and individuation processes. So, to understand cultural identity in digital world as a means to its recognition and preservation we focus on the notion of culture through the relationship between the reception of information [Simondon 1964, 2010] and self organization processes[108] [Debrun 1996a, 1996b, 2009], which – in our hypothesis – are strictly related. We aim to contribute to the explanation of the perception/reception of habits, abilities and beliefs – understood as information – as builders of cultures, here understood as well as a singular knowledge. We are strongly considering individuation processes through language to comprehend cultural processes. It is of our interest to give display to Simondon's philosophy of technology and it's actual relevance to help in the process of recognition and preservation of cultural identity in digital world.

In a broader view this project, which is our attempt to work in this key, aims at the promotion of activities of academic cooperation at national and international levels in Science, Technology and Innovation (STI) towards digital inclusion of K-12 students and teachers, as equally towards the reflection about the conditions of technological innovation in the post-colonial situation among Brazilian researchers as well as researchers of Portuguese speaking countries such as Cape Verde, Angola, Saint Tome and Principe and East Timor. We intend to reflect about the aspects of technological needs of our countries, looking for social, cultural and economic development through a wider circulation of free software among our countries. The goal is to map the possibility of open digital content spread and the broadening of access to information in an interconnected network of free and open digital libraries,

---

[108] *I want to express my gratitude to Prof. Itala D'Ottaviano for inviting me to participate in the Interdisciplinary Group of CLE Self Organization (Grupo Interdisciplinar CLE Auto-Organização), a research group that made me rethink the notions of system and information what allowed to better interact with local communities: http://www.cle.unicamp.br/principal/?q=node/18&language=en.*

and especially as a means to give access to local communities to elaborate and organize culture and knowledge in digital world. Digital libraries recently act as key elements to preservation, dissemination and maintenance of small circulation cultures at international level. In this case, the effort to articulate our experience with other countries permits an interesting comparative view to common reflection about digital inclusion of the so called "minor languages" and "minor cultures".

With the promotion of free and open software in partnership with the University of Waikato (New Zealand), the costs for digital information spread are lower, limited in general to the human resources available to maintenance/creation of information and infrastructure. This proposal is to make viable the use of digital libraries in the study and preservation of cultural heritage in different local languages, as to articulate our debate from the perspective of creating a practice of South-South cooperation and exchange of information in digital space to the production of common information and knowledge.

An academic-cultural approach for projects concerning multilingualism and multiculturalism seemed to be a good way to integrate local cultures, local languages and academic knowledge. So we proposed an Extension project with K-12 students at the Programme for Junior Scientists (Programa de Iniciação Científica Junior – PICJr) to build digital collections in partnership with Comunidade Jongo Dito Ribeiro[109] of Campinas at the "House of Afro Culture Rosebush Farm" (Casa de Cultura Afro Fazenda Roseira), an urban quilombo at the city of Campinas. They wanted to organize in digital collections all the information about their practices, activities, and languages. One of the students has chosen to create a digital collection on Feijoada das Marias do Jongo. This project went so well that we changed the lane of organizing the digital collection to do the registry of the way of doing the feijoada. More than that, the registry of all the necessary steps to do the Feijoada das Marias Jongo developed into the elaboration of a formal process proposed in partnership by the community and the university for the recognition of the "South East Feijoada" as a national intangible heritage. The documents are being reviewed and shall be submitted in November 2014 to the Brazilian National Institute of Historic and Artistic Heritage (Instituto do Patrimônio Histórico e Artístico Nacional – IPHAN). Of course we will have a digital collection about Feijoada Marias do Jongo, much richer than we could have ever thought of. This process also allowed us to better understand the role of African languages in quilombos, such as this one, and to have a glimpse on the existence of African languages that are transmitted vertically

---

[109] *Comunidade Jongo Dito Ribeiro: http://youtu.be/kJz7r57f1oA.*

in the community, that we weren't aware of. So, by all means the preservation of cultural identity is a strong part for understanding multilingualism and multiculturalism in the country, and it is the soul of digital collections.

3) *Example of a Language Processor to Preserve Cultural Identity*

Another interesting project with an original language that would be worth mentioning for preserving cultural identity is the work of Prof. Wilmar D'Angelis, linguist and indigenist at the Unicamp University, who has adapted in 1992 a math software (ChiWriter[110]) to permit the Kaingang script to be written in a digital notepad. At that time word processors did not support the characters of the Kaingang alphabet. It is correct to suppose that word processors will process words, but – as proved by Prof. D'Angelis – it is also correct to sustain that a mathematical software can enable digital writing to local languages. Then Kaingang texts began to be written in digital world, but many other indigenous languages remain outside this media. They are not part of Unicode. And the proper Kaingang remains with the problem of writing directly on the web, action not available in most common software. It is of our interest to discuss such digital inclusions/exclusions, and to underline the creativity and efficiency of this solution.

This discussion, however, is not of our interest only, but concerns indigenous peoples, whose languages are under threat of extinction. We are following with interest the efforts of the project "Web Indígena" (Web-Indigenous), of the NGO Kamuri, for the digital inclusion of the language Kaingang, and Wilmar D'Angelis has sustained important discussions about the question with that ethnic community. Initial results of this project can be viewed on the site www. kanhgag.org.

4. **To Comprehend Linguistic Plurality and Digital Inclusion of Local Languages and Local Cultures in Brazil**

> *It is in my political interest to join forces with those Marxists*
> *who would rescue Marxism from its European provenance*

> Gayatri Spivak, 1998

On multilingual issues, European multilingual reality and European linguistic public policies play a strong role as a model for the current discussion. In a

---

[110] *"ChiWriter was a commercial scientific text editor for MS-DOS, created by Cay Horstmann in 1986. It was one of the first WYSIWYG editors that could write mathematical formulas, even on the very slow IBM PC XT computers that were then common." (http://www.wikipedia.org). "ChiWriter is a scientific word processor that was sold by Horstmann Software Design Corporation between 1986 and 1996" (http://www.horstmann. com/ChiWriter/).*

post-colonial situation, the status of local languages is extremely different. This difference ranges from the value of language within each culture and its role, to its social visibility, or its possible mediatic inclusion and digital portability. Usually, we assume European and USA multilingual standards as academic references, this is one part of the process, the other that I assume would be necessary as well is to discuss the international division of intellectual labor and the international circulation of digital assets. Or, at least, to critically avoid the situation of a reference, an experience abroad, to became a pattern or a goal without considering local history, material context and culture.

The main axes of this reflexion:

A) Post-colonial criticism is an important perspective for this project, considering that the reality of multilingualism is – in the majority of the countries with languages in danger of disappearing and languages of small circulation – due to colonial enterprise, and its particular effects on local cultures. So, we have a common starting point, nevertheless in Portuguese Speaking countries post-colonialism has never been an academic issue. We would be interested in promoting post-colonial archives and debates among the participating higher education institutions, and therefore to enhance South-South theoretical publications in common thematic grounds.

B) Language and culture as necessarily non-established objects. Language and culture play a major role in this debate, and to encapsulate them within a specific knowledge area would be to forbid the appearance of history, memory, unconscious, politics, policies, geography, economics, etc. in our research. The epistemological basis that assures the presence of a language or culture as an abstract and "complete" object usually inflicts the metaphor of variation and/or assimilation to administrate differences, always being related to an established and historically dominant concept of language and/or culture.

Multiculturalism and multilingualism environments here are not dominant patterns, in fact they are most frequently what is left aside the dominant discourse. Also the functions of language and culture must be enabled to vary. We are choosing communicational aspects to deal with information and communication of local languages and cultures. When possible, we would be interested in developing a debate on different approaches of the role of language in societies.

C) Technology as the asset of "world" communication that puts in cause poetry, familiar language, mother tong values, local education quality, local underdeveloped economy, local production of human resources, etc. In this paradoxical comprehension, technology is also the possibility of the impossible. The possibility of dreaming of an equalization on information availability and

accessibility among any existing cultures. Good and continuous education for all. Some countries, some higher education institutions that are our partners, just don't have the minimum necessary to start dreaming together. It is our responsibility to acknowledge this fact, and try to think together on different ways to interact.

To enable our present network to live and to speak, two curious steps are necessary. The first one is to zoom out from the imaginary speed of technological development, and the other is to take a little reflexive distance from European citizenship patterns. Because in this group we don't have either, and we need to start working with what is really available.

References

1. Alencastro, L. F. de (2000). O Trato dos Viventes: formação do Brasil no Atlântico Sul – São Paulo, Companhia das Letras, 525 p. ISBN 85-359-0008-x.

2. Debrun, M. (2009). Identidade Nacional Brasileira e Auto-Organização (Brazilian National Identity and Self-Organization). Coleção CLE, v. 53, 186 p. Organizadores: Itala M. Loffredo D'Ottaviano, M. E. Q.Gonzalez, tradução por Valéria Venturella – Campinas: Unicamp, Centro de Lógica, Epistemologia e História da Ciência. ISSN 0103-3147.

3. Miceli, S. (2001). Intelectuais à Brasileira. São Paulo, Companhia das Letras, 436 p. ISBN 85-359-0113-2.

4. Rodrigues, A. (2002). Línguas Brasileiras: para o conhecimento das línguas indígenas. São Paulo, Edições Loyola, 135 p. ISBN 85-15-01045-3.

5. Simondon, G. (1964). L'individu et sa génèse physico-biologique. Epiméthée, Paris, Presses Universitaires de France, p. 250. In: Laymert Garcia dos Santos... [et al.]. Revolução tecnológica, Internet e socialismo / São Paulo: Editora Fundação Perseu Abramo, 2003. – (Coleção Socialismo em Discussão). ISBN 85-86469-79-3. Outros autores: Bernardo Kucinski, Maria Rita Kehl, Walter Pinheiro.

6. Simondon, G. (2010). Communication et Information: cours et conférences. Les éditions de la transparence/Philosophie, Chatou, France. ISBN 978-2-35051-043-9.

7. Spivak, Gayatri (2006). In Other Worlds. 409 p. Routledge Classic Edition. ISBN10: 0-415-38956-9.

8. Spivak, G. (1999). A critique of postcolonial reason: toward a history of the vanishing present. 449 p. ISBN 0-674-17763-0.

9. Wanderley, C. (2009). A periferia digital e o processo de descolonização. In: África-Brasil. Caminhos da Lingua Portuguesa / organizadores Galves, C.; Garmes, H.; Ribeiro, F. R., editora da Unicamp (colecao Unicamp ano 40), ISBN 978-85-168-0838-6.

10. Wanderley, C. (2010). Bibliotecas Digitais Multilingues: uma proposta inclusiva. In: Greenstone: Un software libre de codigo abierto para la construccion de bibliotecas digitales. Experiencias em America Latina y el Caribe. Editor: UNESCO Montevideo, Gunter Cyranek, UNESCO, ISBN 978-92-9089-149-9

11. Witten, I. H., Bainbridge, D. (2007). A retrospective look at Greenstone: Lessons from the first decade. In: Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries, Vancouver, Canada, pp. 147-156.

Huang CHENGLONG

Research Fellow, Institute of Ethnology and Anthropology,
Chinese Academy of Social Sciences
(Beijing, China)

Chinese Ethnic Languages in Cyberspace

1. Introduction

Since the founding of the PRC, the Chinese government identified the 55 ethnic minorities except for Han Chinese, therefore, China is a unified, multi-ethnic country.

There have been established 155 ethnic autonomous areas, including 5 autonomous regions, 30 autonomous prefectures, 120 autonomous counties since 1946. At the same time, as a supplement of the ethnic autonomous areas, in the minority towns 1173 ethnic autonomous townships have been established. In addition, for 11 small ethnic groups who have no autonomy 9 ethnic minority townships have been established. Just as the book *Languages of China*, edited by Sun Hongkai, et al. [2007] estimates, the 55 ethnic minorities in China use roughly 130 languages, and there are vast differences in the number of speakers and their vitalities.

Sun [2001] made statistics on the language status and population proportion of 128 languages, which is given in the table below.

Number of users of each language	Number of languages (percentage)	Total number of users
below 100	7 (5.5%)	400
101–1000	15 (11.7%)	11,000
1,001–10,000	41 (32.0%)	219,000
10,001–100,000	36 (26.5%)	1,300,000
100,001–1,000,000	17 (13.3%)	12,100,000
1,000,001–10,000,000	10 (7.8%)	31,000,000
Over 10,000,000	2 (1.6%)	1,120,000,000

According to Sun's [2011] preliminary study, 20% of languages are endangered, such as Gelao, Nu, Pumi, and Jinuo. There are 40% of languages on the edge of extinction. Three languages are nearly extinct, such as the Yì language, Mulao

and the Khakas language [Sun 2006]. There are several ethnic languages which are completely out of the communicative function, i.e., Manchu, Nanai (Hezhe), Tatar and the She language.

With globalization and rapid development of information technology, and the era of digital media in the 21st century, the central and provincial governments, scholars and educated ethnic people are working harder than ever for recording, preserving, promoting and disseminating ethnic minority languages via cyberspace.

In this paper, we mainly introduce and discuss the use of minority languages in the Internet in China, such as language teaching and learning, daily conversation, writing system, bilingual dictionaries, folk songs from various websites, including popular Internet portals, mainstream media, language learning and audio-video websites. In the meantime, we explore the reason why the use of minority languages in cyberspace is vastly divergent depending on such issues as geographical region, politics, population, culture, and ethnic self-awareness.

2. Minority Languages in Cyberspace

There is a great difference in how Chinese 55 minority languages disseminate in the Internet. Ethnic languages, which are supported by the central and provincial governments, have a widespread and a great impact on the Web. Some languages are actively supported by local governments and are relatively abundant in language data and online learning materials. Some languages have no support by governments, but they are promoted by volunteers in cyberspace. Several ethnic languages are not present in the Internet. In this section we briefly introduce the promotion and dissemination of radio & audio, TV & video and script of Chinese ethnic languages in cyberspace.

2.1. Internet Radios and Web Audios

Internet radios and web audios mainly refer to the network broadcast and folk songs, language teaching and other audio content.

(1) State Radio Station

The China National Radio (CNR) has China Ethnic Broadcasting Network (http://www.cnrmz.cn/), and there are Mongolian, Tibetan, Uyghur, Kazakh and Korean broadcasts.

(2) Provincial Radio Stations

The Xinjiang Broadcasting Station (http://www.xjbs.com.cn/aod/gushigb/): Uyghur, Kazakh, Mongolian and Kirgiz language broadcasts

The Voice of Tibet of China (http://www.vtibet.com/gb/): broadcasts in Lhasa and Khams Tibetan.

The Qinghai Radio Station (http://www.qhradio.com/) and China Tibet (http://www.tibet.cn/: Tibetan broadcasts.

The Inner Mongolian Broadcasting Network (http://www.nmrb.com.cn/): a Mongolian broadcast.

The Ethnic Minority Channel of the Yunnan Radio Station (http://www.ynradio.com/pinlv/node_125.shtml): Dehong Tai, Xichuangbanna Tai, Lisu, Jingpho (Kachin) and Lahu minority language broadcasts.

The Radio Station of the Chuxiong Yi Autonomous Prefecture of the Yunnan Province: Yi language news.

The Radio Station of Qian South-Western Buyi and Miao Autonomous Prefecture of the Guizhou Province: the Buyi language and the Miao language broadcasts.

The Radio Station of Wenshan Zhuang and Miao Autonomous Prefecture of the Yunnan Province: news in the Miao language and the Yao language.

The Radio Station of the Yanbian Korean Autonomous Prefecture of the Jilin Province: a Korean broadcast.

2.2. Internet TV and Online Videos

(1) National Network TV Station

The China Network TV Station (http://www.cntv.cn/): Mongolian, Tibetan, Uyghur, Kazakh and Korean TV channels.

(2) Provincial TV Stations

The Xinjiang TV Station (http://www.xjtvs.com.cn/): the Uyghur language and the Kazakh language channels.

The China Voice of Tibet (http://www.vtibet.com/tb/sp/zyws/): Tibetan Star TV and TV show.

The Inner Mongolian TV Station (http://www.nmtv.cn/vod/myws/myws.shtml): Mongolian Star TV.

The Yanbian Korean Media Network (http://www.ybrt.cn/): Korean Star TV.

The Xishuangbanna TV network (http://www.bntv.cn/): the Dai (Tai) language and the Hani language news.

The Dehong Government Portal (http://www.dehong.gov.cn/video/): the Dai (Tai) language, Jingpho (Kachin) language and the Zaiwa language news.

2.3. Online Minority Scripts

The China Xinhua News Agency portal (http://www.xinhuanet.com/): Uyghur, Tibetan and Korean script web pages.

The Xinjiang Daily (http://www.xjdaily.com.cn/): Uyghur, Kazakh and Mongolian script web pages.

The China Tibetan News (http://www.chinatibetnews.com/) and the Qinghai News (http://www.hybrb.com/): Tibetan web pages.

The Guangxi Ethnic Newspaper (http://www.gxmzb.net/): a Zhuang script web page.

The Yanbian Daily (http://www.hybrb.com/): a Korean script web page.

The Dehong Union Daily (http://www.dhtjb.com/Html/20105189506-1.html): Dai script, Jingpho script, Lisu script and Zaiwa script web pages.

The Xichuangbanna Daily (http://www.dw12.com/): Dai old script and new script web pages.

The Khams Daily of the Sichuan Province (http://www.kbcmw.com/): a Tibetan script web page.

The government website of the Xinjiang Yili Kazak Autonomous Prefecture (http://www.xjyl.gov.cn/): a Kazakh script web page.

2.4. Minority Languages Databases

Many scholars in different institutions are using Chinese ethnic languages databases, such as the Yunnan minority languages database by the Yunnan Nationalities University, Yunnan Minority Commission's Yunnan minority languages and scripts database, Yi, Qiang and Ersu languages sound databases by the South-Western University for Nationalities. In 2008, the Ministry of Education (MOE) of the People's Republic of China (http://www.moe.edu.cn/) and the State Language Commission of the People's Republic of China (http://www.china-language.gov.cn/) started to establish an *Audio Database of Chinese Ethnic languages*, as those databases have not been completed, and they are not available online.

Nagano (The National Museum of Ethnology) carried out *rGyalrongic languages databases* (http://htq.minpaku.ac.jp/databases/rGyalrong/)[111]. The website stores 425/1200 lexical items and 200 sentences for 81 dialects (including some non-rGyalrongic languages such as Geshezha, Minyag and Lavrung).

3. Concluding Remarks

As shown above, Chinese government gives complete freedom to each ethnic group to disseminate its own language and culture in cyberspace, however, the uses of minority languages online are vastly divergent due to the geographical region, politics, population and ethnic self-awareness.

1. Among the 5 autonomous regions, Xinjiang, Tibet and Inner Mongolia have online media systems of some minority languages, such as radios, TV and websites.

2. There are 33 ethnic groups, who inhabit cross-border territories between China and other countries, and people communicate with each other more frequently in borderline, therefore, their languages disseminate abundantly on the Internet (such as Korean, Kirgiz, Lisu language, etc.).

3. Ethnic groups who possess larger population, higher political position, and more intellectuals (like Yi, Miao and Bai) effectively and systematically disseminate their languages on the Internet.

4. Some ethnic groups are not supported by the central government, but receive support by local governments. Their languages are promoted sporadically by their intellectuals and volunteers in the network communication, that is the case of Dong, Santa (Dongxiang), Qiang and Xibe language.

5. 5 minority languages with no data on the Internet are Jing, Jinuo, De'ang, Chinese Russian and Lhoba (Lopa).

On the 5-6th June, 2014, *The International Conference on Language* (which was organized by the Ministry of Education of The People's Republic of China, the State Language Commission of the People's Republic of China, National Commission of the People's Republic of China for UNESCO, Jiangsu Provincial Government) was held in the Suzhou municipal, Jiangsu Province

---

[111] *This is part of the research results of the MEXT / JSPS overseas research grant A-21241007 "International joint survey of the rGyalrongic languages", 2009–2012 fiscal years, edited by Yasuhiko Nagano and Marielle Prins.*

of China. More than 100 countries participated and 400 people attended the conference. Objectives of the conference were:

- To address the importance of language ability for intercultural understanding and sustainable social development.

- To enhance language ability through the exchange of information, good practices and knowledge on innovative approaches to language education, including multilingual education.

- To promote collaboration amongst concerned stakeholders across different regions to enhance language ability.

The conference reached consensus: "The efforts of ethnic and indigenous populations to transmit their languages across generations are crucial for a more just and productive world. <...> Cyberspace should reflect the linguistic diversity of the world and all language communities should benefit from the potential of information and communication technology (ICT)."

Language policies and practices responding to the needs of national, indigenous and immigrant communities can enhance effective communication for peaceful co-existence in the global society.

As we know, language is the most important part of human culture, and language diversity is the basis of cultural diversity. However, in some countries language is not regarded as a part of intangible cultural heritage, which is extremely disadvantageous for documentation and conservation of endangered languages and for maintaining language diversity. We suggest that UNESCO clearly states the place of language in intangible cultural heritage, and then secure an international convention to establish a set of persistent and effective measures for promoting language diversity.

References

1. Huang, Chenglong (2009). Qiangyu Fangyan Tuyu jiqi Huoli (Qiang subdialects and their vitalities), In: Zhang, Xi, Huang, Chenglong and Lan, Guangsheng (eds.), *Chidian Fuwei: Qiangzu Wenhua Zaihou Chongjian Xingsi (The Rescue of the Qiang People: Introspection on the Post-earthquake Reconstruction of Qiang Culture)*, 177–197. Beijing: Minzu University of China Press.

2. Huang, Chenglong, Li, Yunbing and Wang, Feng (2011). Jilu Yuyanxue—Yimen Xinxing Jiaocha Xueke (Documentary linguistics – A new interdiscipline). *Yuyan Kexue (Linguistic Sciences)* Vol. 10 (2011).3: 259–269.

3. Sun, Hongkai (2001). Guanyu Binwei Yuyan Wenti (On the endangered languages in China). *Yuyan Jiaoxue yu Yanjiu (Journal of Language Teaching and Research)* 2001.1: 1–7.

4. Sun, Hongkai (2006). Zhongguo Binwei Shaoshu Minzu Yuyan de Qiangjiu yu Baohu (The salvage and protection of the endangered minority languages of China). *Jinan Daxue Xuebao (Journal of Jinan University)* 2006.5: 126–129.

5. Sun, Hongkai (2011). Yuyan Binwei yu Feiwuzhi Wenhua Baohu (Language endangerment and protection of intangible cultural heritage). *Yunan Shifan Daxue Xuebao (Journal of Yunnan Normal University)* Vol. 43.2: 1–7.

6. Sun, Hongkai, Hu, Zengyi and Huang, Xing (eds.) (2007). *Zhongguo de Yuyan (The Languages of China)*. Beijing: The Commercial Press.

Turrance NANDASARA

Director, National Online Distance Education Service,
Ministry of Higher Education of Sri Lanka;
Senior Lecturer, University of Colombo School of Computing
(Colombo, Sri Lanka)

Yoshiki MIKAMI

Director, Language Observatory;
Professor, Nagaoka University of Technology
(Nagaoka, Japan)

Bridging the Digital Divide in Sri Lanka: Some Challenges and Opportunities

Abstract

The "digital divide" is a gap in technology usage and access. It has been investigated by scholars[112] and policy makers[113] mainly as an economy-specific issue that permeates the population across all demographic profiles, such as income, gender, age, education, race, and region, but not specific to the languages of different communities. The lack of native language driven ICT is a major conducive factor in the digital divide.

The Sinhala writing system used in Sri Lanka is a syllabic writing system derived from *Brahmi* and consists of vowels, consonants, diacritical marks and special symbols constructs. Several of these constructs are combined to form complex ligatures. The total number of different glyphs in the Sinhala language is close to almost 2300. Thus, all computer equipment that supports this language needs to support also a greater degree of complexity in both displaying and printing with near minimal changes to the keyboard or input systems. A case of "Digital Inclusion" described in this article shows how small communities of non-Roman script users can connect to the Romanized-system-dominated cyberspace.

---

[112] *Hoffman, Donna L.; Novak, Thomas P. (1998). Bridging the Digital Divide: The Impact of Race on Computer Access and Internet Use. Educational Resources Information Centre, Department of Educatopn, U.S.A.*

[113] *NTIA (1999). Falling Through the Net: Defining the Digital Divide, National Telecommunications and Information Administration, U.S. Department of Commerce, U.S.A.*

1. Introduction

In Sri Lanka, which has a population of 21 million people, the majority are Sinhalese (74.9%). Other ethnic groups include Sri Lankan Tamils and Indian Tamils (15.4%), Sri Lankan Moors (9.2%), Malays, and Burghers.[114]

Primarily, there are three living languages in Sri Lanka. They are Sinhala, Tamil and English, used for general, every day communication: both interpersonal and mass communication. Of them, Sinhala and Tamil are considered "national languages" while English is considered as a "link language" connecting major ethnic groups of the island. Thus, written documents, on paper or other materials, appear in one, two or all of these languages.

The Sinhala language has a syllabic alphabet in which all consonants have an inherent vowel /a/. This alphabet differs from all other Indo-Aryan languages: it contains special sounds that are unique to it since the 8th century A.D.

1. There is a unique set of five nasal sounds known as "half nasal" or "prenasalised stops" in the Sinhala writing (*ňga, ňja, ňḍa, ňda* and *mba*). These five consonants have no equivalents in any of Indic languages.

2. A pair of unique vowel symbols (*æ* and *ǣ*) to represent two vowel sounds is in use since the 9th century A.D.

This article focuses on the key issues and the structure of the Sinhala writing at character level. Then it progresses to examine the design and development of scripts for deferent technological generation such as printing and typewriters, followed by major steps behind the design of the ISO-10646 code. We will also show how this was used to try to bridge the digital divide in Sri Lanka.

2. Sinhala Scripts and Major Features

Sinhala is a uniquely spoken and written native language in Sri Lanka. The Sinhala script is used for writing the Sinhala language. Sinhala is said to have derivatives from the ancient scripts *Brahmi*, known to have existed since the 3rd to the 2nd century B.C.E. Subsequently the alphabet and writing systems have changed considerably with notable influence by the *Kadamba* and *Pallawa Grantha* scripts of south India[115]. The full Sinhala script includes symbols

---

[114] *The 2015 World Factbook (2015). US, Central Intelligence Agency.*

[115] *Florian Coulman The Writing Systems of the World, Blackwell Publishers Ltd., 1989; Florian Coulman, The Blackwell Encyclopedia of Writing Systems, Blackwell Publishers Ltd., 1996; Fernando, P. E. E., Palaeographical Development of the Brahmi Script in Ceylon from the 3rd Century B.C. to the 7th Century A.D., University of Ceylon Review 7, 1949.*

necessary for writing *loaned* words, that is, words that originated from Sanskrit and Pali, notably those with aspirated consonants.

2.1. Sinhala Scripts

The modern Sinhala alphabet comprises letters of the *Eḷu*[116] alphabet and the Sanskrit alphabet. It contains 61 letters, of which 18 are vowels (V), 41 consonants (C), 2 diacritical marks (D), 16 vowel signs (V) and (X), and 2 special symbols (Cry). This alphabet is used in writing *Eḷu*, Pali, Sanskrit, and foreign words naturalized in the language.

2.2. Major Features of the Writing System

In the Sinhala language, combinations of consonants, vowel signs and semi-consonants produce different phonetic sounds.

Some consonants and vowel signs are combined to form syllable blocks (glyphs). Some syllable blocks are "unpronounceable" and are not used in the written system or spoken Sinhala. Some glyphs are constructed in a different way according to the letter shape. Some would create a rather uneven, irregular and illogical outer appearance.

Every combination is constructed according to the shape of the Sinhala letters. Forty one (41) consonants (C) and sixteen (18) vowel signs (V) can be combined to form a glyph. Thereafter, each united glyph can be further combined with 2 special symbols (Cry), *rakaransaya* and *yansaya*, and then even further it can be combined with 2 diacritical marks (D). After all it will produce more than 2300 "usable" combinations for the Sinhala writing.

However, these combinations are more complicated when single or multiple vowel signs are attached to the same character. Keeping in view the major issues outlined in this section related to graphical representation of character composition and combinations in the Sinhala writing, it would be interesting to see how this language developed over a fast two and a half centuries until computer machine came into operation.

---

[116] *The term Eḷu is given to the pure dialect of Sinhala unmixed with foreign words, and Siṅhala to the mixed dialect, though in point of signification the two terms have not the least difference. Sihaḷa in Pali, Siṅhala in Sanskrit and heḷa in Eḷu*

3. Historical Development of the Sinhala Writing System

3.1. Background

The oldest writing of Sinhala can be traced back to about the 3rd century B.C. These are inscriptions mainly marked by either cave or rock[117] found in almost all parts of the island. Usually these cave inscriptions are found below the drip-ledge where the script too is protected from water (Figure 1). In some cases, the writing continues as one line for about forty to fifty feet from left to right and in some cases it goes from right to left.

Figure 1. *The Vessagiri Cave Inscription of the 2nd century B.C.*
(Source: Author's collection, 1998)

3.1.1. The Golden Era of Ola manuscripts

The Sri Lanka Museum in Colombo has a collection of about 3600 ola leaf manuscripts. The oldest palm leaf manuscripts in existence are the *Dhampiyā Aṭuvā geṭapadaya* (belongs to the 10th century), *Chūla Vagga* (12th century), *Amāvatura* (12th century), *Saddharmaratnāvaliya* and *Pujāvaliya* (13th century). From the 13th century A.D. onwards, production of literature became more prolific[118].

3.1.2. Era of Type Printing

Remarkable progress was made in documentation systems after the arrival of the Portuguese in Sri Lanka in 1505. The Portuguese compiled a list of records

---

[117] *The University of Cambridge, England has 274 volumes of 'Epigraphica Zeylanica' with over 3000 inscriptions from Sri Lanka (that is more inscriptions than the whole of mainland China has), including one dating back to the 6th century B.C. Over 2000 of these have been deciphered, indicating the consistent development of the Sinhalese language.*

[118] *Piyadasa T. G., Libraries in Sri Lanka, Their Origin and History from Ancient Times to the Present Time, Sri Satguru Publication, India, 1985, pp. 1-18.*

(*tombōs*) of villages to aid with tax collections. In 1658, the Dutch became the masters of the costal districts for the period until 1796). The Dutch maintained those records and also made a more important contribution of charting the area on maps. Dutch settlers started schools for Europeans and also for local people. In these schools, the language of medium was their own mother tongue. Seminaries were established in northern Sri Lanka for higher education, where Sinhala and Tamil were taught as special subjects. Hence, such educational activities demanded books in these languages.

The first book of any size ever printed in Sinhala was a 41-page Sinhala prayer book published in 1737 by *Gabriel Schade* (*see* Figure 2).

Figure 2. *A page form the first book printed in Sri Lanka (1737)*

(Source: Department of National Archives, Government of Sri Lanka, Original image was scaled down to 40%)

In 1876, The Royal Print Shop in Vienna, Germany, printed a comprehensive samples register *"Alfabet des Gesammten Erdkreises"* [Alphabet of all races of the world]; it consisted of 29 sheets[119]. A printing example of 76 languages recorded in this sample register, among which was Sinhala, has been recorded as Cingalesisch (German for Sinhala). Such printing was indicative of the European interest in Sinhala scripts (Figure 3).

[119] *Alfabete des gesammten Erdkreises aus der K.K. Hof- und Staatsdruckerei in Wien, 1876.*

Figure 3. Sinhala printing types from the *Alfabete des Gesammten Erdkreises*, 1876
(Original image was scaled down to 75%)

3.1.3. Printing Establishments in Sri Lanka

The first Sinhala newspaper in Sri Lanka was "*Lakmini Pahana*" which commenced publication on the 17th September, 1862. It is seen that the whole newspaper was printed using only one type face and one size of type. Even the headlines were set using the font used for the text. The Sinhala printing was well established in Sri Lanka by the mid-19th century.

3.1.4. Sinhala Monotype Type Faces

Major improvement to the quality of typeface design and layout came with the arrival of a new mechanical typesetting machine, the monotype. A Monotype Sinhalese font Series No. 557 (*Sinhalese*), No. 657 (*Sinhalese Bold*), No. 698 (*Sinhalese Italic*) and No. 699 (*Sinhalese Bold Italic*) consisted of 302 characters, 26 punctuation marks, and numbers in each font type[120].

120 *Monotype, Book of Information. New ed. London, 1959.*

4. Major Steps in Sinhala Text Processing

4.1. The First Sinhala on Computer Screen

With the introduction of BBC microcomputers to the University of Colombo in 1982, the staff of the university's Statistical Unit and one of the authors, S. T. Nandasara, pioneered the development of a set of Sinhala Bitmap fonts for computers. Using this Sinhala font set, daily TV programme schedule was transmitted to the public by the national TV station Independent Television Network (ITN), and it was the first ever known attempt in Sri Lanka to use computers with local-language capabilities.

4.2. Long Debates on Alphabet Related Issues

Since the mid 1980s, the Sri Lankan government has taken steps to address Sinhala-language-related discrepancies, one of which was use of different alphabetical orders by different dictionaries[121]. As a result, some of the committees involved have been working on a Sinhala-language character set, on essential features and shape of each character, its alphabetical order, and thereafter, on a standard keyboard and a standard code for the character set during 1985–1998[122]. A new character to denote *fa* was introduced formally to the standard (SLASCII-Sri Lanka Standard Sinhala Character Code for Information Interchange) as the last character in the set.

4.3. A Sinhala Font Set for Early Computers

The Institute of Computer Technology (ICT) of the University of Colombo initiated the incorporation of Sinhala capabilities into personal computers in

---

[121] *Working Paper, 1985. Order of Alphabet and System of Transliteration, CANLIT & NARESA.*

[122] *Samaranayake, V. K., Disanayaka, J. B. and Nandasara, S. T., 1989. A standard Code for Sinhala Characters, Proceedings, 9th Annual Sessions of the Computer Society of Sri Lanka, Colombo; Nandasara, S. T., Disanayaka, J. B., Samaranayake, V. K., Seneviratne, E. K., and Koannantakool, T., 1990. Draft Standard for the Use of Sinhala in Computer Technology, Approved by the CINTEC on the advice of its working committee for recommending Standards for the Use of Sinhala and Tamil Script in Computer Technology; Nandasara S. T., 1990. "Working Group Paper, Draft Sri Lankan Standard proposal for Sinhala Character Code for Information Interchange", Working Group for Sinhala Code for Information Interchange, Coordinated and Managed by Computer and Information Technology Council of Sri Lanka; Samaranayake, V. K., and Nandasara, S. T., 1990. A Standard Code for Information Interchange in Sinhalese, ISO-IEC JTC1/ SCL/WG2 N673; Nandasara, S. T., 1991. "Proposed Sri Lankan Sinhala Standard Code for Information Interchange (SALASCII)", Approved for the Computer and Information Technology Council of Sri Lanka (CINTEC) and Submitted to the Sri Lanka Standards Institute; Samaranayake, V.K. Nandasara, S.T., 1994. A Standard Code for Information Interchange in Sinhala, Proceedings of the International Conference on the Input and Output of National Character Sets, AFSIT-8, Tokyo, Japan.*

1998. At a later stage, capabilities were incorporated to maintain the Tamil language character set, diacritical marks, mathematical and phonetic symbols for DOS operating system[123]. Language was selected by toggling the Shift-Ctrl key combination wherever required.

4.4. SLS 1134:1996 – A Phonetic Model

The new, phonetic model design for the Sinhala character code replaced the older typewriter metaphor concept from the previous SLASCII standard. ISO/IEC's formulation of the Sinhala Unicode Standard was based on proposals submitted by the early contributor from Ireland[124], the United Kingdom[125], the USA[126], Japan[127] and Sri Lankan[128]. After a few ad hoc committee meetings, national delegates and other nominated country delegates decided to accept the character set, names, and arrangement for the Sinhala script[129] based on the Sri Lankan proposal with slight modifications, and the Sinhala Code Chart was included in the Unicode Version 3.0[130]. The SLS 1134:1996 standard and the Sinhala keyboard layout were also modified appropriately[131].

---

[123] *Nandasara, S. T., 1997. Sri Lanka Experience of Development of Tamil Input/Output/Display Methods, TAMILNET'97 – International Symposium, Singapore; Nandasara, S. T., Leong, K. Y., Samaranayake, V. K., and Tan, T. W., 1997. Trilingual Sinhala Tamil English National Web Site of Sri Lanka, INET97, http://www.isoc.org/inet97/proceedings/EI/E1_3.HTM; Nandasara, S. T., Samaranayake, V. K., 1997. Current Development of Sinhala/Tamil/English Trilingual Processing in Sri Lanka, MLIT-2, November 7-8, Tokyo, Japan.*

[124] *Michael Everson, 1989. Proposal for encoding the Sinhala script in ISO/IEC 10646 (revision 1). http://www.evertype.com/standards/si/si.html; Michael Everson, 1996. Report of the Sinhala Standards, ISO/IEC JTC1/SC2/WG2 N1473R.*

[125] *Hugh McGregor Ross, 1996. Sinhala proposal, ISO/IEC JTC1/SC2/WG2 N1376.*

[126] *Lloyd Anderson, Ken Whistler, Peter Lofting, Rick McGowan (Contributors), 1992. Draft Proposals Unicode Technical Report #2 Unicode Inc.*

[127] *Naito Eisuke, 1998. Progress Report of the MLIT Project, AFSIT-12, Ha Noi, Vietnam; Takayuki K. Sato, 1998. Status of Cooperative Activities for the Missing Characters, MLIT Secretariat, CICC, Japan.*

[128] *Samaranayake, V. K., and Nandasara, S. T. 1990. A Standard Code for Information Interchange in Sinhalese, ISO-IEC JTC1/SCL/WG2 N673; SLS 1134:1996. Sri Lanka Standard SLS 1134:1996-Sinhala Character Code for Information Interchange, SLSI publication.*

[129] *Michael Everson, Takayuki Sato, Kohji Shibano, Disanayaka J. B., Nandasara S. T., Johan van Wingen & Glenn Adams, 1997. Minutes of Sinhala Ad-Hoc Committee, Doc. # 1613, ISO Meeting No. 33, Iraklion, Creete, Greece.*

[130] *The Unicode Standard 3.0, 1998. (www.unicode.org), Addison-Wesley Pub Co., ISBN 02001616335*

[131] *SLS 1134:2004. Sri Lanka Standard SLS 1134:2004-Sinhala Character Code for Information Interchange, SLSI publication.*

4.5. *Current Development Platform Status*

The government of Sri Lanka began to focus on IT issues in the mid-1980s. After Unicode had been established in 1998, unlike Thailand, India and Nepal where governments work closely to develop local versions of Microsoft OS, in Sri Lanka this issue was not on the agenda until the recent years. To overcome this problem, government must have solid policy and broad active plans to invest on local language development not only in their own soil, but also continuing long term lineup work agreements with such corporations.

In addition, a non-standard keyboard input technology is being developed to allow users to insert, download, search, and create new data and build their own content in the Sinhala language. Using roman text to represent Sinhala words and sentences (for sending SMS and in online chat rooms) is widely spread. Since the demand increases, this may encourage local information/Internet service providers to design proper and standard technology to put up quality information in all three languages so that Internet technology can take root in Sri Lanka. It is to be noted that development of such technologies is not in the ISPs general agenda.

5. The Research Infrastructure

The local language research infrastructure in Sri Lanka is very much geared towards short-term objectives. The academics felt they could use direct Masters level students in short three month projects only. As soon as the students were at a level where they might do something useful, they graduated and left, leaving the supervisors to start the whole process once again. This situation is compounded by Sri Lankans doing their PhDs abroad where they are invariably steered away from working on issues related to local languages. Their research field will be selected according to the western interest. Almost all the Sri Lankans who did their PhDs in Western, European, Scandinavian or even Eastern countries had not worked on the localization field. As a consequence longer term investment in large scale projects in IT research is almost unknown until recently. There are aid and grants for longer term investment in other areas – for example the British, Swedish, Norway and Japanese Government funds are for long-term investment in environmental research, infrastructure development, human resource development, but not for research related to local languages and even not for IT related research. In 2004, the IDRC's localization initiatives and funding helped the University of Colombo's School of Computing to establish a Local Language Research Centre. Private industry has banded together to create the Software Vendors Association, but, while having research funds available, it focuses on issues such as local area networks and other basic items of commercial infrastructure. Language processing is not on their priority list.

6. The Digital Divide

Though the Internet is gradually evolving as a mass media around the world, introducing information technology in Sri Lanka is not mass oriented, unlike other media such as news papers, books, radio, television, etc. It is limited in orientation, and is also government-oriented. Though much is said, less of that is actually delivered to the end user. A lot needs to be done, and the Sinhala language is yet to be approached by technology initiatives. However, localization approaches alone may not be successful enough to bridge the digital divide. It has to be discussed not only from a linguistic perspective, but also in the context of other technical issues, since it involves convergence of many related technologies, economic resources (per capita) available to people to take steps to cross the digital divide, and so on.

Let us remember that everything that English has, or for that matter, what Sinhala is going to have, cannot be or may not be appropriated by other languages used in Sri Lanka, i.e. Tamil, Pali, Sanskrit, etc., but they certainly can increase their vitality by becoming part of the IT world in as many possible ways as they can. At present, however, most of them are out of race.

7. Conclusion

Language identity and its recognition have to be maintained. This can be done through local radio or TV stations. It could also be done on a low cost technology[132] and world wide scale with the help of UNESCO, an organization committed to preserving human culture and languages and to narrowing down the digital language divide[133]. Education is another way of promoting and preserving languages as the means of digital divide. How should we best teach and provide learning opportunities for human languages? Building multilingual word dictionaries, maintaining common social context and common oral corpus would be useful for multilingual communities. Thus, creation of digitized corpora is a basic task in the attempt to preserve the world's languages. Corpora can be multimodal, spoken or written, depending on the type of linguistic material and recording equipment is available. This will bring to a discussion of the role of language technology once again. Probably a key factor here is reuse of technology for similar languages. This will bring the solution as mentioned in the 'E-commerce and Development Report 2003' by the United Nations

[132] *Wijayananda Jayaweera, Kothmale Community Radio/Internet Project: Expanding the Knowledge Base, UNESCO, June, 2001.*

[133] *Nandasara, S. T., GII/GIS for Equal Language Opportunity, MLIT-3: October 6-7, Ha Noi, Vietnam, 1998.*

Conference on Trade and Development (UNCTAD), providing an insight into what software developing countries can use for bridging the digital divide.

Acknowledgements

References

1. Beeching, W. A. (1974). *Century of Typewriter*, Heinemann, London.

2. Daniel, J. T. K. & Hedlund R. E. (1993). *Carey's Obligation and India's Renaissance*, Council of Serampore College, Seranpore, West Bengali.

3. De Silva, M., Sugathapala, W. (1982.) *Some Consequences of Diglossia.*

4. Lankage, J. (1996). *Evolution of Sinhala Scripts*, S. Godage & Pvt Ltd, Colombo, SriLanka.

5. Miller, E. (1883). *Ancient Inscriptions in Ceylon*, PLATES, Trubner &Co. Ludgate Hill, London.

6. Paranavitana, S. (1970). *Inscriptions of Ceylon*, Vol. I.

Valerii DIOZU
Member of the Board, Chisinau Municipal Council
(Chisinau, Republic of Moldova)

Linguistic Preferences in the Moldavian National Cyberspace: Reflection of Political, Economic and Migration Processes in the Society

First and foremost, I would like to express deep respect to the true enthusiast and patriot of Russia Mr. Evgeny Kuzmin and gratitude for this opportunity to admire the remarkable places of your great country and to share my opinion on the issues of multilingualism in cyberspace.

In this paper, I'm going to analyze the linguistic preferences in the cyberspace of the Republic of Moldova that reflect the country's political, economic and migration processes. This is an attempt to seize on the relationships and understand the principles underlying the ongoing processes in the local information environment.

Forty years have passed from the date when two computers were linked successfully and a distributed computer network prototype was created to become a basis of the modern **World Wide Web.** We have to admit the big success of that project – the rate of its development has exceeded all the technological solutions known to the humanity, including radio, television and telephony. Today, the Internet is deeply integrated in our routine life, and such basic principles as domestic communication and individual freedom are losing an unequal contest with it. The WWW allows us to easily find people with similar interests and approaches to life as well as old acquaintants who, for some reasons, have dropped off our radars. It is psychologically easier to start communication on the Internet than face-to-face. All these advantages have laid a basis of Internet communities which have been playing a significant role in the life of our society.

Today, few people are involved in the analysis of the essence and layout of the World Wide Web. Regulatory bodies, computer networks, computers and protocols are not in the focus of the end users' interests, and this state-of-the-art has become a norm of our routine life. People are dreaming of getting access to the vast arrays of structured data which accumulate the wisdom of the ages. We hope that, thanks to the Web and the diversity of information technologies, each of us will get the long awaited access to the rich storages of human knowledge which will be unrestricted both quantitatively and geographically. Based on this knowledge we will make an unprecedented leap in our development – the leap which will overcome the borders of the

information society and will lead us into the depth of cyberspace. However, those of us who are involved in these processes are envisaging many problems and shifts that are hindering this movement.

In view of a vast number of theories that are used for describing the processes within the information environment and of the absence of clear principles and criteria, I've tried to limit the issue under study and focus on the national cyberspace of the Republic of Moldova.

As soon as we start considering a national cyberspace we come across numerous restrictions, challenges and paradoxes that block the road to a happy information future. These factors are less obvious in the global Internet society. Let me name some of them:

1. It is impossible to speak of any physical or logical restriction for a national information environment (restricted by the state borders or national domains).

2. Preservation of the linguistic borders of a national information environment goes along with the total exclusion of monolinguicity because the linguistic capabilities of any member of the society increase the volume of the national environment and shape its multilinguicity.

3. For some historical reasons (the birthplace of the Internet and the use of English as the language of technology), predominance of English content in the WWW discriminates non-English communities in access to information.

I'd like to draw your attention to a new task of analyzing the psychoactivity of the Internet and its effect on the immature mind of children who immerse into this environment from infancy.

The number of these challenges is gradually increasing as a result of the development of cyberspace and application of new technologies. As early as 1997, Friedman and Nissenbaum made an attempt to determine and classify the types of these challenges and shifts. The situation looks desperate.

However, I presume that it is this dynamism that carries our salvation. Many of the challenges are temporary. Technological progress and reasonable regulatory policy will help us deal with these challenges and move beyond. An example of this is the development of machine translation technologies which make content in foreign languages generally accessible. This technology resolves the challenge of information access discrimination of non-English-speaking communities.

In other words, we will always face the task of continuous monitoring, analysis and control of the WWW development for the purpose of giving fast response to the current challenges and shifts.

All this may be equally applied to national cyberspaces.

To describe the object of the analysis, let me introduce the following criteria:

- The availability of the governmental body or a national environment.
- The availability of the advances or technologically developed infrastructure.
- The availability of regulatory rules.

Brief Info

Moldova is located in the south-western part of the Eastern European plain on the territory of 33,700 sq.km.

As of January 1, 2012, the population of the Republic of Moldova was 3,559,500 people. As of the census of 2004, 75.8% of the total population was Moldavians; 8.4% – Ukrainians; 5.9% – Russians; 4.4% – Gagauz; 2.2% – Romanians; 1.9% – Bulgarians; 1% – other nationalities.

General Understanding of the Moldavian Cyberspace

According to the Statistics Department, Moldova has 1,439,000 households, of which 78% have at least one computer. 730,000 are connected to the Internet. The number of subdomain names in the MD segment exceeds 25,000.

According to the National Regulatory Agency for Electronic Communications and Information Technologies, it is possible to see the growing percentage of the number of users of mobile and fix Internet to the total number of users.
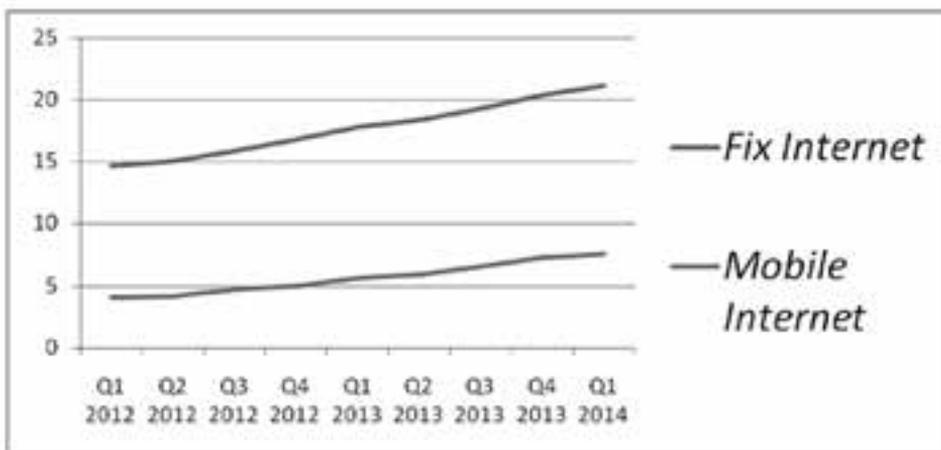
According to the NetIndex's rating, Moldova leads in the Internet rate as provided by the Internet service providers. By the download rate, Moldova is placed 17 with the rate of 17.5 Mbit/sec; but by the upload rate, Moldova is placed 7 with 10.5 Mbit/sec.

According to the Circulation and Internet Audit Bureau (BATI), Moldova's Internet environment has 1,502,637 unique users.

ICT contribution to the country's GDP equals 10%. Thus, the share of the information sector in the national economy is quite significant.

Based on the information from Alexa, 100 of the most active web-sites of the MD domain area have been selected and divided into the following groups:

1. Governmental – 12%.
2. Noncommercial – 23%.
3. Commercial – 36%.
4. News – 16%.
5. Others – 13%.

For 18 months we have been monitoring these web-sites on the availability of linguistic pages. The results are shown below.

Governmental web-sites
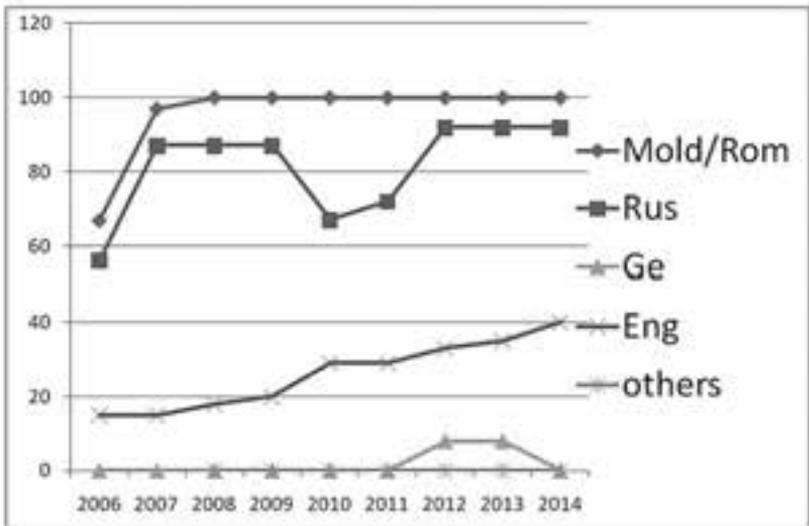
*Ge – web-pages in Gagauz.

News web-sites
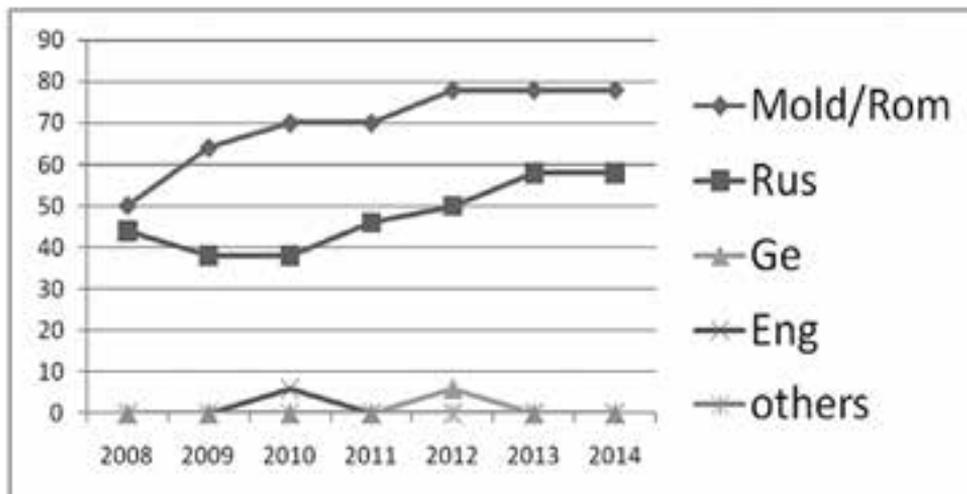
Commercial web-sites
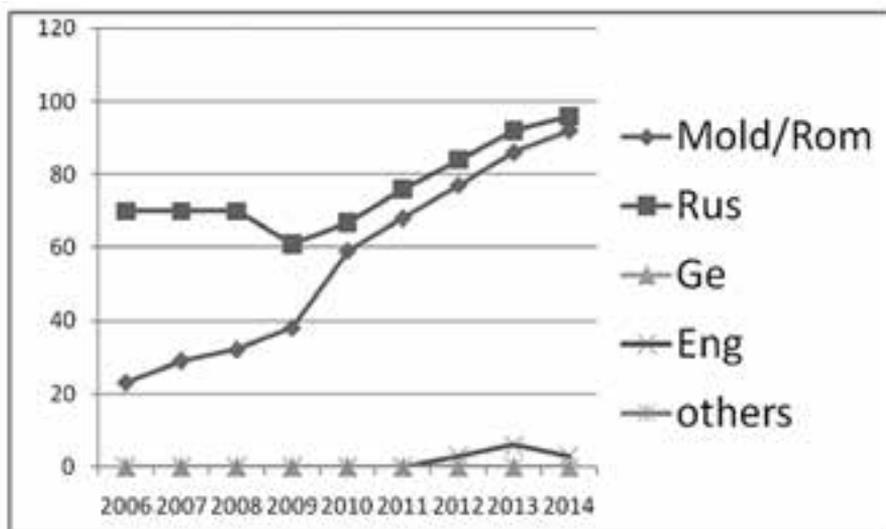
Noncommercial web-sites
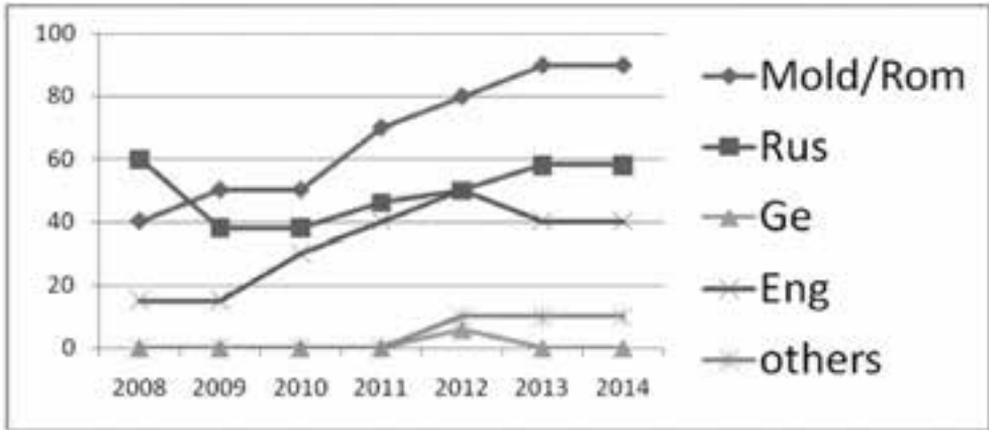
Noteworthy is that all critical points of the graphs coincide with the decision-making periods in both political and social spheres (2009 was the year of changes in the country's political vector; 2012 was the year of elections to the People's Assembly of the Gagauz Autonomous Region; 2013 was the year of signing the European Union Association Agreement).

The graph below shows the number of unique users of Moldavian web-sites according to the query initiating country.
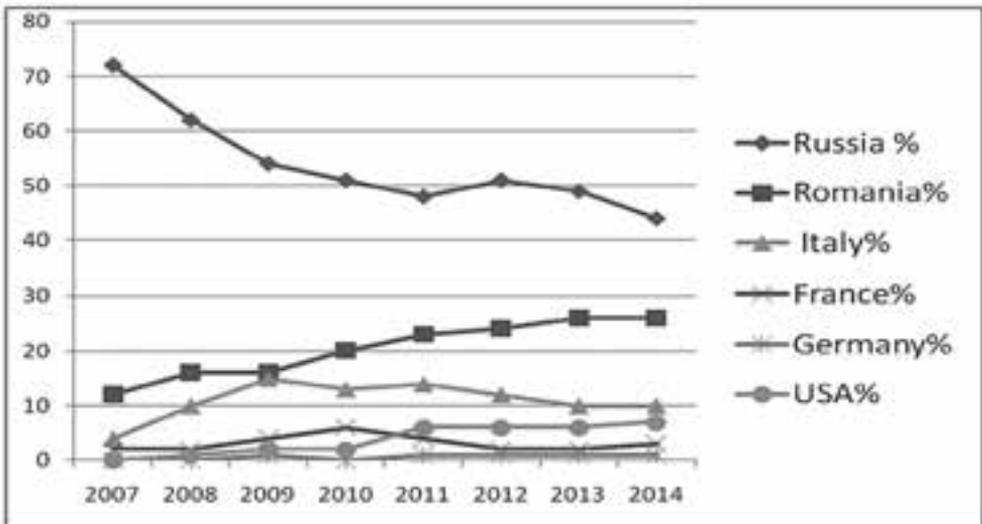
This graph is in surprisingly close correlation with Moldova's recent official migration flows statistics.

Taking in consideration the above, we can forecast the possibility of creating in the near future the approaches to control and management of social migration and economic processes in Moldova by way of analyzing various layers of cyberspace, creating individual social monitoring information systems. However, as I have already mentioned, we should be always aware of the emerging challenges and shifts.

Anuradha KANNIGANTI
Lecturer,
National Institute of Oriental Languages and Civilizations (France)
(Hyderabad, India)

Defending Languages in India: A Socio-Economic View

Abstract

I raise some questions regarding the limits of a "patrimonial" and "identitary" view of language defense, which appears to underlie many recent expressions of anxiety concerning cultural loss and attempts to "protect" languages, addressing myself in particular to the Indian context.

I propose that the focus must be shifted to exploring substantive responses to the powerful socio-economic factors driving the decried symptoms of language shift even concerning major languages, rather than merely responses of "cultural protection" which increasingly fall on deaf ears among the general population driven more by considerations of economic mobility than language loyalty.

In the interest of finding such responses, more research energies should be devoted to the political economy of language.

In this direction, I propose "vernacular (language) economies" as an object of study, with practical implications for how they can be developed, and thus for the maintenance of linguistic diversity.

Introduction

The 3rd International Conference on Linguistic and Cultural Diversity in Cyberspace has as its declared objectives the creation and implementation of policies to develop and preserve languages, through the identification of appropriate ethno-linguistic policies for stopping or slowing down the processes leading to language marginalization; to help languages to be "better equipped, represented and used"; and to explore policies, standards and tools towards these ends.

Among the key prerequisites identified are policies for "securing languages' presence and development in cyberspace".

In my contribution, I propose to raise the less-evoked question of linguistic diversity in the economic terrain and language policies in relation to economic development. As in the case of cyberspace, what policies might be identified for

"securing languages' presence and development in the economy"? I refer here more to the problematics of the formal economy linked to formal education systems, particularly in developing nations on rapid growth trajectories, as the high stakes terrain of marginalization or realignment of native languages faced with economically dominant and/or global languages – "It's the economy, stupid", describes well the nature of the forces of marginalization.

A related question is: should language valorisation and defense be conceptualized solely viewing languages as identity markers and carriers of "culture"? This appears to be both a popular understanding and the tenor of the discourse of language defense. What are the consequences of such a partial understanding, when in fact millions are making language choices that rhyme more with personal utility maximization?

Language Defense and Economics: A Disconnect?

Mufwene and Vigouroux point out the divergence between advocates of language preservation who celebrate linguistic diversity, and those who reflect on and devise economic policy, who see a multiplicity of languages as an obstacle to economic development:

> *"(L)anguage advocates have tended to conceive of languages as cultural representational systems, which must be maintained (regardless of whether they are adaptive to the ambient socioeconomic ecologies)..."* [Mufwene and Vigouroux 2014].

Language activists consider that privileging one or a handful of languages for the sake of economic development, in particular promoting an education system in which everyone learns the economically dominant language, leads to language endangerment, disadvantaging those who have to give up their heritage languages in the "linguistic marketplace" – but, while the possibility of being schooled in one's "own" language may offer a better learning experience, it may also render the student ultimately less competitive in the lucrative modern economy. In effect, languages can count both as assets and liabilities for their speakers.

> *"(O)ne cannot deny that language shift... has often helped particular people adapt (better) to socioeconomic ecologies that would otherwise be disadvantageous to them".* [Mufwene and Vigouroux 2014].

On the other hand economists viewing the reduction of linguistic diversity in the economic terrain as advantageous for populations rather ignore the importance of ethno-linguistic identity, of languages as social identity markers and carriers of culture-specific world views.

However it cannot be denied that the trends of economic integration on global and national levels that push for the adoption of dominant national or global languages in education and the formal economy ("English only"), have been a leading cause of language marginalization or even language shift. This trend is not correlated to the size of the language community – leave alone "small" languages, even very "large" ones can be marginalized in the domain of the formal national economy, as in the case of major Indian languages faced with English.

Even in the case of economically salient if not "global" languages, such as many national languages in Europe, marginalization trends have to be examined, with domain-specific "language shifts" happening. A case in point is the progressive adoption of English as the language of higher technical education in certain European countries, the national language being henceforth reserved for arts, humanities and social sciences [ICEF Monitor 2012]. This trend, motivated both by European integration considerations and global ambitions, points up the threat to the preservation of linguistic diversity in domain knowledge, knowledge production and economic creativity and innovation.

Market Value of Multilinguism

Oustinoff [2012] highlights a counter-trend to marginalization with the emergence of the "language industry". He points out that, while economists did not take a serious interest in language until the late 20th century, its globalization "made it impossible to ignore language. From the economic point of view, languages became a considerable and strategic industry in their own right." [Oustinoff 2012: 409].

This new economic angle to language focuses on the key element of reaching out to markets, involving language learning, translation, marketing in local languages and product localization. From the angle of global business, multilinguism goes beyond being merely a cost that needs to be minimised:

> *"Once considered a question of secondary importance, the linguistic potential of nations now constitutes a major strategic asset in the era of economic globalization. ... (This) translates into no less than a major paradigm shift, which doesn't consider languages as expendable, but lends them major market value, not only in terms of cost but in terms of profitable investment."* [Oustinoff 2012: 416].

In essence, this view is about considering speakers and communities as consumers and markets respectively in the globalised market economy: "the realization that it's in the language of the Other that selling must take place". The shift in perspective is not only due to the difficulties of generalizing a

putative cost-minimising "English-only" policy as the means to facilitate global communication (language acquisition being a slower process than the spread of markets) but is also due to the emergence of a multi-polar world which highlights language as soft power.

However, in spite of the recognition of the "market value" of multilinguism, in reality, the basket of languages that are admitted to be "worth their while" in terms of market value is quite limited. For example, the number of students enrolling each year to learn Hindi in France's National Institute of Oriental Languages and Civilisations is quite significant (it has been known to go up to 150 enrollments), but it's not related to any perceived market value of Hindi in the global economy (though it's the fourth most spoken language in the world), rather the motivations are largely cultural ("Bollywood"). This does point to the "cultural dimension" of the language economy, but as Grin [2002] admits, this is "non-market value". Grin highlights the difficulties of developing the concept of languages as value:

> *"(I)t is not possible, for the time being, to truly calculate (i) the value of a language, (ii) the "value" of one linguistic area in relation to another, (iii) the "benefits" (market and non-market) to expect from a particular policy; (iv) many of the costs, direct and indirect, associated with such an initiative..."* [Grin 2002:21].

Even if we were to find a way to synthesize the different values (Grin identifies four of them) into one measure, the overall value would likely reveal that a limited number of the world's languages actually make it into the league of the economically desirable.

Political Economy of Language

In approaching the economic angles to language, beyond the conceptual difficulties, at the outset there appears to be an issue of definitions and nomenclature, between the "Economy of Languages" or "Language economy" as in the citations above, and "Economics of Language", Language Economics, or "Language and the economy", as revealed in the programme of a recent international conference in Paris [Mufwene and Vigouroux 2014].

I propose to add to these emerging formulations, and emerging objects of study, the themes of "language(s) in the economy", the "language part of the economy" (inspired by "the language part of work", [Boutet 2001]) and the "economy of a language".

Beyond the traditional focus of language planning on language in education, media and administration, and the newly salient concept of the "market value

of languages" in the globalization phase that views speakers and communities as consumers and markets respectively, I propose a shift in focus to consider them as producers and communities of production respectively, within emerging growth trajectories of national economies. In addition to the "market value of a language", the idea is to explore concepts in relation to "language as a tool of production", and "the productive scope of a language" within a given economy.

A tentative approach to the "economy of a language" is to view it as "the totality of productive activities performed in that language, that can be measured in terms of the wealth or value created by such activities". "The value of a linguistic area" mentioned by Grin [2002] might be pertinent in this context, as well as the notion of "minority economy" discussed in [Lukanovic 2014].

I shall specifically employ the term "vernacular economy(ies)", in speaking of post-colonial national contexts such as India, where the colonial language (CL) has the dominant role in the formal economy, in which the indigenous languages tend to be marginalized. Thus, one could speak of the "economy of French" in Senegal, as opposed to the "vernacular economies" of Wolof, Diola and so on; the "economy of English" in India, as opposed to the "economy of Hindi" or the "economy of Telugu".

In approaching this, I'm inspired in part by studies of the political economy of religion and caste in India, such as those of Harris-White [2003], where she explores differential accumulation of economic assets and employment patterns among religious communities and among caste groups. Comparable studies of political economy could be made in the case of the language factor, which would of course be more complex given that language is not exclusive unlike religion or caste, and does not necessarily define "community" in a delineated manner.

The other part of inspiration is a conviction, born of observation of the Indian language scene, that enhancing the "economic vitality" of languages (a concept needing elaboration) is essential to their survival, in a context of rising economic aspirations. Note that this is not the same as the "economic vitality of language communities", as referred to in [Beaudin 1996]. Tools to describe the character and size of a vernacular economy could serve to estimate and track the economic vitality of the language, identify spaces of opportunity and factors of blockage, and thus contribute to policy, planning and investment in the vital triangle of language, education and livelihoods development.

Policy: Language, Education, Economy

These observations raise key questions concerning the policies to be adopted for "securing languages' presence and development in the economy." What kind of economic development, within what kind of "socio-economic ecology" and which sectors would enable the usage of economically marginalised languages and enhance their economic vitality in a national space? How can this be reconciled with the drive to economic integration for which the less number of languages of economic activity, the better? When can one position be better than the other, as a function of context?

> *"Do languages really have rights to education and economic systems in the same ways that citizens have rights to these institutions", and if so, how shall such rights be financed?* [Mufwene and Vigouroux 2014].

Mufwene [2014] calls this is a "wicked problem" to which one has perhaps to be contented with a "satisficing" solution – of a multilinguism privileging the use of fewer languages than may be desired for the ideal of linguistic diversity. Such a solution would arise from a particular reconciliation of the two conceptions of language, as a representation system and as a communicative tool, within particular national contexts – population structures, configurations of education, training, livelihoods and mobility patterns, the phase of economic development and, not least, communication practices and political constructions of language – all giving rise to particular kinds of competition between languages.

In the following, I consider the case of India – its configuration of languages in relation to the economy, coming from a particular history, that has coalesced over the last two decades into an accelerated marginalization of Indian languages with respect to English, not only in economic domains but spilling over into cultural transmission [Tully 2011, Hariharan 2104].

The reaction of the anglicizing classes to the trends of cultural marginalization is an interesting reflection of the diglossia at work – having consciously made their language choice, and confronted with language shift, they are taken up by great cultural anxiety and fall back on the conception of language-as-representation-system.

Language and Economy in India

In post-Independence India, access to English was both the preserve of the entrenched elite and the factor ensuring its reproduction, while the discourse of democratization focused on the use of "vernacular" languages (VL) in education and administration for the masses. (While census data shows that India has 122

languages with at least 10,000 native speakers, I restrict myself here to the 22 official languages which concern over 95% of the Indian population – however I use the word "vernacular" to emphasize their marginalised status in the formal economy). There is a two-tier educational system inherited from the colonial era: private "English medium" for the elites, and public "Vernacular medium" schooling for the masses. In English medium schooling, the local VL may be studied as a "subject" (language of culture); in Vernacular medium schools, English is a second language. This structural element neatly crystallises the duality "language of mobility" and "language of culture".

Notwithstanding nationalist sentiment about the languages after Independence and the reorganization of the nation into states largely defined according to language, the "market value" of the VLs has always been limited by the almost total preference for English in the modern economic sectors and to a large extent even in government. English has been set up as the pre-eminent (and exclusive, at least in popular perception) language of socio-economic mobility, with the exception of some success stories of vernacular medium students (for the most part, trajectories of ascension to the English economy). No veritable alternative narrative has emerged, but the recent arrival of Narendra Modi as Prime Minister of India, a complete product of vernacular education with little acculturation to the Anglo-sphere, has opened up a new space of assertion from "vernacular" India, changing the idea that *"only the poorly educated, or the provincial-minded, or those from the lower classes preferred to speak an Indian language instead of English."* [Subramanian 2104].

As has been pointed out, the English/Vernacular duality defines a linguistic 'caste system' of its own, and the Anglicized elite have never been interested in the "language question". In the rapid growth era of globalization and the market economy, the massive rise of "aspirational" classes has led to an accelerated and generalised clamour for access to English (it having turned out to be a huge competitive advantage for India in outsourcing and other service sectors).

> *"India's aggregate human development has been neglected in favor of the success of the elite, who have global aspirations. This is why the elite holds on to English and why the rest of India aspires to it."* [Pillalamarri 2014]

It can be argued that it's the IT/ITES (IT enabled services) connection, the model of socio-economic ascension that it represents, which has single-handedly swept away any remnants of defense of "mother tongue" education, and has made the desirability of a single programme agenda of "English For All" (let alone "Education for All") an eminently political question, with the underlying assumption that English equals a job in the IT sector and a H1 visa to the U.S.A. down the line. So much so that access to "English medium" has in fact become

a part of populist electoral promises. The clamour for English as the fast track to mobility has even taken on a "caste" colouring, with certain caste formations setting up the language as a "Goddess" ("The English Goddess") who will deliver them from their disadvantaged status, by elevating them to that other "high caste", the English-knowing caste, as Nehru called it [Devraj 2010].

But the fact is that the IT/ITES sector can assure a few million jobs at most – it cannot accommodate the vast population of the aspiring young whether they manage to master English or not, let alone the intricacies of effective teaching of English to ever vaster numbers – the failures of English for All are well documented, with students acquiring neither their native language nor English.

> *"How many jobs actually need knowledge of English in order to function? Not many. India's growth cannot be powered by the service industry and call-centres alone, many of which are saturated anyhow. A very minute percentage of Indians will work in such industries and those that do can learn English as a skill for their job, which would ultimately be more efficient than trying to educate a large portion of the population in the language."* [Pillalamani 2014]

Beyond this IT fixation, the elephant in the china shop is the belief that "English is necessary in the context of globalization" – but necessary for what, and for who? What is the nature and requirement of language use in different domains and at different scales? These questions have not been elucidated.

The Language Conundrum and the Economy: Global vs. Local Languages

A very instructive case in point that highlights the confused tussle over English is the recent debate in India over its weight in civil service exams [Rangan 2014]. Those arguing for English insist that future administrators will need to communicate on a pan-Indian scale – whereas the large part of an official's work tends to be in local contexts where the VL is of primary value. In fact as Akurathi points, VL candidates, rather than those schooled in elite English language institutions, are the best adapted to working in local administration, being closest to its realities.

> *"Ironically, all that effort to learn English, and the humiliation faced along the way, seems ridiculous now. A majority of our daily transactions are in Kannada. I would be a miserable administrator if I were proficient in English but didn't speak Kannada"* [Akurathi 2014].

It's not recognized that a similar fallacy operates in the reasoning about language in post-secondary education, when English is taken to be indispensable for

318

participating in the modern economy. A typical view on this is the following, reflecting the very prevalent dualistic understanding of language referred to above.

> *"Cultural chauvinism is misplaced in today's world. That is not to say that we should ignore our language and culture. Preserving our language and culture is important, but so is ensuring our economic competitiveness."* [Mazumdar 2014]

This confuses competitiveness in the global economy with the local components of economic activity, both informal and formal, which realistically concern the larger part of the population and we may surmise a major chunk of the national economy. The tangible role of VLs in economic activity, and the potential for their development within a veritable modern "vernacular economy" does not get the attention it deserves under the steamroller of English-fixation, reinforcing the poor image of the languages as having little economic utility, or at least not the desirable type with the aspired for socio-economic status.

The result is that a particular equation has been settled between language, education and livelihoods: the acquisition of the dominant language of national and global mobility gets fixed and generalised as "the" means of socio-economic success, and this shapes language choice in education down to socio-economically modest classes, where we observe a large-scale rush for "English medium" even on the part of the rural poor, ready even to sell their modest assets to pay for private schooling.

The Defense of Language: The Problem of a "Patrimonial" View

I wish to argue that the "culturalist" conception of language defense ties into and reinforces the confusion over language and economy between the local and global levels . This view of cultural loss and the idea of protection that has been increasingly expressed by different actors concerned by the changing linguistic equations in Indian society are typified by the above citation of a leading Indian industrialist [Mazumdar 2014]. It illustrates the problematic conception of language defense in the context of a "diglossic" configuration assigning distinct crystallised economic and cultural functions to a "global" as opposed to a "local" language. The linguistic duality of "English-Vernacular" underlying the two-track educational system in India generates cultural, social and economic dualities and fractures, in turn shaping a duality of discourses on language loss and defense.

An interesting terrain for examining the response to the marginalization of the vernaculars are festivals for "Defense of Language and Culture". Different state governments in India generously finance "World Conferences" on the

state language, with the declared objective of finding strategies to protect language, culture and identity in the context of globalization. Particular concern is expressed in relation to children and youth, on convincing them that "just because English is necessary in the context of globalization, it doesn't mean their native language should be put aside". The declared project is to provide the young with the tools "to be competitive" and yet prepare them to simultaneously "protect their language and culture". (World Telugu Conference, Wikipedia)

It's instructive to contrast the relationship of the two groups to the "discourse of loss" – the anglicized and anglicizing (products of English medium), as against the vernaculars.

In the first place, the alarmist discourse about the abandonment of the "mother tongue" by younger generations is most often expressed by those coming from precisely the class of people who have consciously and pointedly pursued the acquisition of English for their children, to position them favorably in Indian society and in the global marketplace. "Culture" and "mother tongue" appear to become important for this class once the advantageous socio-economic position made possible by English, is assured. They are then alarmed when their children, for whom they "purchased" the right tone and the right accent of English in the right schools (while drilling into them subtly or not so subtly, the value of this exclusivity), take the process to its logical end, by a de facto language shift to English.

The model put forward as a response to this undesirable side-effect is in effect to engineer a "diglossic" child, a delicate feat of ideological fashioning in which the child must view English only as a language of advantage in the marketplace, while conscientiously keeping the "mother tongue" as the language of culture.

> *"In these days of globalisation parents may send their children to English-medium schools but must encourage their children to speak in Telugu at home."* [New Indian Express 2012]

This quixotic project of neatly compartmentalizing the languages ("Pursue English for advantage, keep the mother tongue for culture") was to all appearances assimilated by the state governments, according to the press releases concerning for example, the last World Telugu Conference in December 2012, in the state of Andhra Pradesh in south India. Scholars and specialists were to figure out how to devise such a configuration, how to bring the children who've been pushed to acquire as much Englishness as possible, back to Telugu and "Teluguness".

320

In contrast to the culture-losing English-medium educated class within India, are the vast majority who are the products of "vernacular" medium education. Proponents of "mother tongue education" have cried themselves hoarse about the integrated quality of a child's educational experience and personal development when there is no cultural and linguistic rupture between home and school. So it can hardly be anyone's case that these vernacularly-defined young people, unlike those of "English medium", are "losing language and culture". Given their structurally sub-altern position, the "vernacular" parents are hardly losing any sleep over cultural loss, having other, more substantial, worries. They would sell their skin to get their wards out of the "vernacular" trap of limited opportunity, and vast numbers of them are doing so.

The juxtaposition of the cultural loss anxiety of the English-anointed advantaged classes with the socio-economic conundrum of the disadvantaged vernacular-educated, throws up a curious complementarity of pre-occupations which neatly sums up a cultural fracture around language in India, and brings us to the crux of the question. The strategy of the linguistically disadvantaged to bridge the gap is to rush pell-mell to buy whatever variant of English they can get, at the price they can pay – but the buyers of English, especially the "cheap" variant, don't realize that they could end up as the losers on both ends, as pointed out above. This is a danger of far greater proportions than the angst of the advantaged classes over cultural loss.

> *"Not agreeing with the apprehension that Telugu … was one of the endangered languages in the world, (Mr.Nikhileswar) … wanted the government to enforce use of Telugu in all official matters …(It is) people of labour, farming and other poor communities (who) are still speaking in their native idiom and are unknowingly contributing to protection and development of the language."* [New Indian Express 2012]

But these left-out classes (the majority) are not the reference of the state-sponsored festivals of local languages, when they propose to "recognize, celebrate, practice, protect, encourage and promote" language and identity. Indeed the crucial social problem which should be the state's pre-eminent concern is that several dozens if not hundreds of millions of young people are misled into thinking that a difficultly accessible and expensive pursuit of English (no matter what promises are made at election time) is a guarantee of access to the global marketplace, and that only such access is a gage of socio-economic opportunity and success.

Response to Economic Marginalization: An Industrial Culture in the Vernacular

It is indeed possible, and essential, for "social justice", and cultural vitality in the long run, to create economic opportunity for the "vernaculars", by addressing gaps in the language-education-livelihoods equation. In the context of an expanding economy like India's, many new economic spaces emerge that need infusions of skilled manpower and entrepreneurship. India has a massive "skills gap", identified by Indian industry as well as the Indian government, which have recognized its crucial role in the economic growth. This skills gap induced in large part by the almost exclusive focus on university education at the tertiary level (assured in English), and the neglect of TVET (Technical and Vocational Education and Training). Even those advocating for higher education in the vernaculars tend to aim at "degrees", and to paraphrase Alice, "they have to do all the running they can do, to *keep* in the same *place*. If they want to get somewhere else, they must *run* at least twice as fast", to keep up with English. On the other hand, masses of future workers and entrepreneurs of the industrial economy can be trained in targeted VL, rather than purveying the false allure of "English and IT for All", with a calibration of language policy in education and training, according to the projected needs of local, regional, national and global economic spaces. Industrial actors can be convinced to invest in the technical development of the chosen VL to serve as industrial languages.

Those with the capacity, whether vernacular or English medium, may indeed aim for the global marketplace. But vernacular-trained skilled workers should be able to make a good living without the pretense of becoming globe-trotting "techies". When welders and masons are able to pass into the middle class without having to buy English, the problem of language or culture loss can only recede – rather, we can very well have a massive social and cultural shift in the positive sense. What might be the impact on growth of the entry of these empowered "vernacular" actors in the economy, as technicians, skilled factory workers, entrepreneurs... with some requisite knowledge of English?

Governments concerned about defense of culture should be pre-occupied with these considerations concerning the "vernacular" majority of the population, and less with the cultural apprehensions of the "globalised" sections of the present generation. Sheer numbers at least, place the future of language and culture in the hands of the former.

Conclusion

I've discussed an economic angle on linguistic diversity, that entails moving beyond considering speakers as members of ethno-linguistic communities with "culture" as the focus, or as citizens or subjects of governance, to viewing them as economic actors, that is as producers and consumers.

I've argued for moving beyond the trope of "defense" of language – for sufficiently large languages – to language development and expansion of opportunity, particularly in contexts of growth. This would be a positive programme, while "language defense" comes across often as quite an uphill proposition.

An understanding of vernacular economies and their investment and growth potentials as well as perspectives at different levels of a national economy would militate in particular against the fallacious idea that global language(s) are indispensable for socio-economic mobility across the board. In richly multilingual nations, "satisficing" solutions to the issue of allocation of investments and language infrastructure among competing languages would have to be found.

To conclude, for the assured survival and enrichment of a language, it should not be "branded" just only as a language of "culture", but also developed as a language of production and of economic weight.

References

1. Akurathi, P. (2014). A battle I didn't need. Outlook, August 18. http://www.outlookindia.com/article/A-Battle-I-Didnt-Need/291639.

2. Beaudin, M. (1996). The Socio-Economic Vitality of Official-Language Communities. New Canadian perspectives.

3. Boutet, J. (2001). La part langagière du travail: bilan et evolution. In: Langage et Société, 4, no.98.

4. Devraj, R. (2010). 'Goddess of English' breaks caste chains, Asia Times Online, Nov 17, http://www.atimes.com/atimes/South_Asia/LK17Df02.html.

5. Grin, F. L'économie de la langue et de l'éducation dans la politique de l'enseignement des langues. Strasbourg: Language Policy Division, Council of Europe.

6. Hariharan, V. (2014). Indian languages under threat in the digital age. Hindustan Times, June 17. http://www.hindustantimes.com/comment/analysis/indian-languages-under-threat-in-the-digital-age/article1-1230559.aspx.

7. Harris-White, B. (2003). India Working: Essays on Society and Economy. Cambridge University Press.

8. ICEF Monitor. Trend Alert: English spreads as teaching language in universities worldwide. http://monitor.icef.com/2012/07/trend-alert-english-spreads-as-teaching-language-in-universities-worldwide/.

9. Mazumdar, K. (2104). The Language Conundrum. My Thoughts and Expressions, 7 May. http://kiranmazumdarshaw.blogspot.in/2014/05/the-language-conundrum.html.

10. Mufwene, S. S. (2014). Language and Economy: A wicked subject matter. In: Economy and Language – An Inter-Disciplinary Workshop, June 2014, Paris.

11. Mufwene S. S. and Vigouroux C. B. (2014). Concept Paper, Economy and Language – An Inter-Disciplinary Workshop, June 2014, Paris.

12. National Skill Development Agency, Government of India, 2009. http://labour.nic.in/upload/uploadfiles/files/Policies/NationalSkillDevelopmentPolicyMar09.pdf.

13. New Indian Express (2012). Concrete action needed to protect Telugu: Nikhileswar, 21 December 2012. http://www.newindianexpress.com/states/andhra_pradesh/article1388297.ece.

14. Novak Lukanovic, S. (2014). The value of mastering languages in economy. In: Economy and Language – An Inter-Disciplinary Workshop, June 2014, Paris.

15. Oustinoff, M. (2012). The Economy of Languages. In: Vanini, L., Le Crosnier, H. (eds.). Net.lang. Towards the Multilingual Cyberspace. Caen: C&F Editions.

16. Pillalamarri, A. (2014). Why India must move beyond English. The Diplomat, July 11. http://thediplomat.com/2014/07/why-india-must-move-beyond-english/.

17. Rangan, P. (2014). Brown Sahib's Club? Outlook, August 18. http://www.outlookindia.com/article/Brown-Sahibs-Club/291637.

18. Subramanian S. (2014). India After English? The New York Review of Books, June 9.

19. Tully, M. (2011). Will English kill off India's languages? BBC News, 29 November. http://www.bbc.co.uk/news/world-asia-15635553.

Miguel PALACIO
Lecturer, SS Cyril and Methodius Theological
Institute of Post-Graduate and Doctoral Studies
(Moscow, Russian Federation)

Languages of Colombia's Indians:
Current State and Role in the Cultural Life of Colombia

Colombia is a most dynamically developing country of Latin America. Located on the territory of over one million square kilometers in the northwestern part of South America, it is placed third by the economy and fourth by the population in the region. It owes its name to Christopher Columbus who never visited this land. As is well known, his ships reached the New World in 1492. This event went down into history as the Discovery of America but Americans prefer to call it "The Meeting of Two Worlds".

The latter name is justified by the fact that the term "discovery" refers to uninhabited territories while North and South Americas had been inhabited by tens of millions of people by the time of the arrival of Spanish conquistadors. Noteworthy is that the indigenous people of Yakutia, whose hospitality we enjoy these days, look like Indians, and it is in no way by accident. Tens of thousands of years ago, the people of the modern Asian part of Russia could freely travel to the northern part of the New World across the so-called Bering Land Bridge which had connected the two continents before it went under water at the end of the Ice Age. So, the Yakut can easily be called distant relatives of Native Americans.

A most mysterious autochthonous culture of not only Colombia but America in general is San Augustin (in the south of Colombia). The remains of this culture are represented by numerous sculptures of human-jaguars and other anthropomorphic creatures but not by the evidences of any human beings. As a result, one may get an impression that the residents of San Augustin merely dissolved in time and space along with their movable and immovable property. Later Colombia's cultures, such as Chibcha-Muisca and Tairona, are known mainly through their golden masterpieces whose elegance and refinement are envied even by the jewelers of the 21st century. The golden pieces of the indigenous people are exhibited in the Gold Museum of Bogota, the most visited museum collection of Colombia.

The legend of the Golden Land, Eldorado, stems from pre-Spanish Colombia. According to the ancient tradition, a new Chibcha-Muisca Chief led the induction

ceremony at the sacred Lake Guatavita, 50 km from Bogota (the then Bakata, the site of the Chibcha capital). During the ceremony, the Chief, who was covered by golden sand and accompanied by priests, reached the middle of the lake on the float board and jumped into the water from it, while the priests threw golden things into the lake. That was the sacrifice to the goddess of water. When the Spanish came to the Chibcha in the early 17th century, they were told this legend and called the site Eldorado. The conquerors set a goal of draining Guatavita in order to find the treasure of the Indians, but all attempts they made failed.

In the present-day Colombia there live 1,392,623 people who are attributed to Indians. They constitute 3.4% of the total population of 45 million people. There are 82 Indian ethnic groups in Colombia. They are united within 4,141 communities and live on the territory of 34 million ha, i.e. 30% of the national territory (which equals the territory of Finland or Vietnam).

Colombia is culturally and ethnically inhomogeneous. Since the country is crossed by the Cordilleras, the regions were developing independently, according to their respective traditions and mentality. For instance, the seaside of the Caribbean Sea has always been the centre of trade and predominance of liberalism. Central Colombia, on the contrary, has been mostly conservative, and its people, the residents of Bogota predominantly, are known as reserved and even arrogant people. These differences have survived to this day. Geography has seriously impacted the historical development of the country.

In the second half of the 20th century, Colombia faced the problems of guerilla fighting and drug trafficking. The autochthonous population became the most vulnerable social segment during this national confrontation. The guerilla camps and drug laboratories are located in hard-to-reach selva parts which are inhabited by indigenous tribes; both confronting parties often get rid of the Indians as undesirable eye-witnesses to their crimes.

Interestingly, the plant, which has become a source of wealth for drug dealers and guerrillas, is cherished by the Indians as God. I am speaking about the cocaine plant, whose leaves, when mixed with chemical substances, transform into the drug most hazardous for human physical and mental health. The Indians venerate Mother Coca as the Goddess capable of protecting them from hardships and enemies. When the Europeans started conquering the land of Latin America, their natives turned to the God of the Sun who told them to believe in coca which could help them but would turn the Europeans insane and make them lose shape when they discovered it. This is exactly the effect of cocaine. However, the plant itself, according to the Indians and some medical professionals, has therapeutic value and may be used as a stimulator for various physical functions. It was especially useful for the natives in the colonial period when they were turned into slaves and servants by Spanish hidalgos.

Colombia is among the countries with the richest linguistic diversity. Colombia's Indians speak 65 languages that pertain to the Macrochibcha, Jê-Caribbean and Andean-Equatorial macrofamilies. The most widely used Indian languages of the country are Chibcha (Bogota region) and Zenú (Atlantic coast). Out of these 65, over 50,000 people speak three languages; 10,000–50,000 people speak eight languages; 5,000–10,000 people speak nine languages; 1,000–5,000 people speak 11 languages, and less than 1,000 people speak 34 languages.

The Constitution of Colombia recognizes Indian languages as official in their respective habitats and warrants the indigenous people the right to primary education in two languages, Spanish and native. In 2010, the Law 1381 on indigenous languages was adopted. It declares the need to preserve and protect indigenous languages, i.e. all Indian languages plus two Creole languages of Afro-Colombians and Palenquero, the language of the residents of some Caribbean regions of Colombia.

The Ministry of Culture of Colombia has initiated a series of sociolinguistic studies for the purpose of determining the vitality of Indian languages. The results obtained are deplorable. The languages that were used by the Pre-Columbian jewelers to pass the tricks of the trade are quickly disappearing due to the assimilation of the Indians and lack of governmental attention. The Indian-oriented policy of the Colombian government is rather weak in general. President Juan Manuel Santos pays occasional visits to indigenous tribes. Prior to his first inauguration in 2010, he had even passed the spiritual inauguration ceremony with the people of Kogi, Wiwa and Arhuaco in the Sierra Nevada de Santa Marta natural reserve. However, these campaigns are of a truly populist nature.

Note that Colombian scientists pay great attention to native languages of the country. There are research centres in Bogota and other cities (some of them are on the premises of major universities such as the National University, the Andean University and the University of Antiokia). The establishment of the Indian University is underway (though slowly enough) in the Cauca Department. Teaching in this University will be done in indigenous languages. Ethnic educational centres have been opened in some "Indian" regions of Colombia.

Recently, the Colombian Indianistics has made some significant steps forward, and each step has acquired at least one expert in the linguistic community. Symbolically, among them are ethnic Indians who managed to make a scientific or teaching career. The engagement of indigenous peoples into social life is different. In general, they can freely contact with other residents of the country and tourists. The most socialized are those Indians who live in the frequently visited national parks. They produce and sell souvenirs, guide tours and inform visitors of the life and traditions of their peoples. The least visible in the Colombian society are Amazonian tribes who prefer to lead an isolated life.

And what can be said about the Indians in the literature of the country? Known throughout the world is the book «One Hundred Years of Solitude» (1967) by the Nobel award winner Gabriel Garcia Marques. Many stories in this amazing book, which was written in the genre of magic realism so popular in the 20th-century literature of Latin America, are an artistic representation of Indian legends of the Caribbean Colombia, the birthplace of the recently deceased author. Magic realism like no other genre reflects the vision of living typical of the indigenous peoples.

A masterpiece of the Colombian literature is the Yurupari legend, an epic of the Amazonian tribes of Colombia and Brazil. This epic had been passed from mouth to mouth until the late 19th century when it was written in the Nhengatu language by the Brazilian Indian Maximiliano Jose Roberto. Nhengatu is spoken by some 8,000 residents of the Amazonian tribes of Colombia and Brazil. This is a kind of an international communication language since it serves not only intercommunity contacts but also communication between the indigenous tribes and non-Indians. The original Yurupari text had been lost and has reached us in translation which was published in 1890 by an Italian geographer and photographer Count Ermanno Stradelli. The Yurupari legend contains unique data on how Indians visualize the world and their historical route, on the rules of community life of the indigenous tribes. This makes it similar to Olonkho – the central epic of Yakutia whose culture may be, in my opinion, perfectly described using the genre of magic realism. Noteworthy is that the mythology of the peoples of Siberia and the Russian Far East has many parallels with the mythology of the New World.

However, we are talking about the nonmaterial heritage of Colombia prior to Columbus. What is the place of the ancestors of ancient tribes in the present-day life of the country? Unfortunately, 3.4% of the population are regarded by most of Colombians, to say nothing of foreigners, as a living history, a relic of the mysteriously beautiful past. If asked to pronounce a couple of worlds in any Indian language, an urban Colombian dweller will smile and will be at least surprised.

Today, Colombia is on the threshold of probably the most significant event in its history since the victory in the 1821 Spanish American War of Independence. For the last three years Havana has been the site of quite successful peace negotiations between the government of Santos and the leading guerilla organization known as Revolutionary Armed Forces of Colombia. Like never before, the country has come close to the end of the 50-year conflict that has, directly or indirectly, touched every Columbian family. Hopefully, the establishment of the long-awaited peace will contribute to closer attention of the society to the burning needs which have been pushed to the sidelines. Among these needs are preservation and promotion of the Indian culture in all respects.

Nikolai PAVLOV
Administrator of Sakha Wikipedia
(Yakutsk, Russian Federation)

Wiki-Projects in the Regional Languages of Russia[134]: Two Development Scenarios

Abstract

Under conditions of resource deficit, the people in charge of crowdsourcing linguistic cultural projects in the regional languages of Russia should set the most exact development strategies for their projects. These strategies may depend on both subjective and objective factors – the former being the ultimate objective, and the latter being the state-of-the-art and relevancy of the language.

Web 2.0 Languages: General Features

Globalization has given rise to nonlinear, superstrata development of minority languages – the development which is characterized (and determined to a large extent) by more media forms of expression which replaced the so-called naturally usual, linear development. Such development is explained by the adoption of the "western" consumer standard economy by many countries and is closely connected with the development of the global society in compliance with the globalized model. These changes in the society force minority languages to change in favor of several languages dominating in the world, while losing on this way their respective grammar features in phonetics, vocabulary (first and foremost, nouns), word formation and phraseology (the latter two suffer from direct borrowings), leaving syntax in peace for a while (though we are witnessing some novelties here as well).

From the standpoint of functionality, it is more reasonable to group the Web 2.0 languages[135] not by their common root but to divide them by the influencing language. Mind that division of minority languages by the fields of influence on contact languages is gradually becoming more important. In practice, this means that regional languages of the ethnic groups of Russia, which we are going to discuss later, find themselves in the field of strong influence of Russian

---

[134] *Regional languages of Russia are languages of the indigenous peoples of Russia that differ from the national official language, Russian. Sometimes, this definition is continued by adding the words "and that do not have the national status outside the Russian Federation."*

[135] *Web 2.0 is a technologically developed community in which communication through social networks plays the dominating role.*

and will, for some time, remain in this field irrespective of the language family. However, the influence of English is noticeable even today, though it is indirect as yet. With time, due to deepening economic and social integration between language communities, other influences will become inevitable: from English, in the direct form that time, and also, from oriental languages (most probably, from Chinese to the largest extent) following the tightening economic and social relations with Asian countries.

These tendencies should be taken into account while developing new language projects and assessing their prospects.

Wiki-Projects in the Regional Languages of Russia

Wiki-projects are, in essence, web-sites on CMS[136] MediaWiki. Historically, the first web-site of this kind was Wikipedia – an online universal encyclopedia. Later, it became multilingual and was joined by a multitude of wikipedias[137] and other projects of Wikimedia Foundation, such as Wikiteca, Wikinews, Wikidictionary and others.
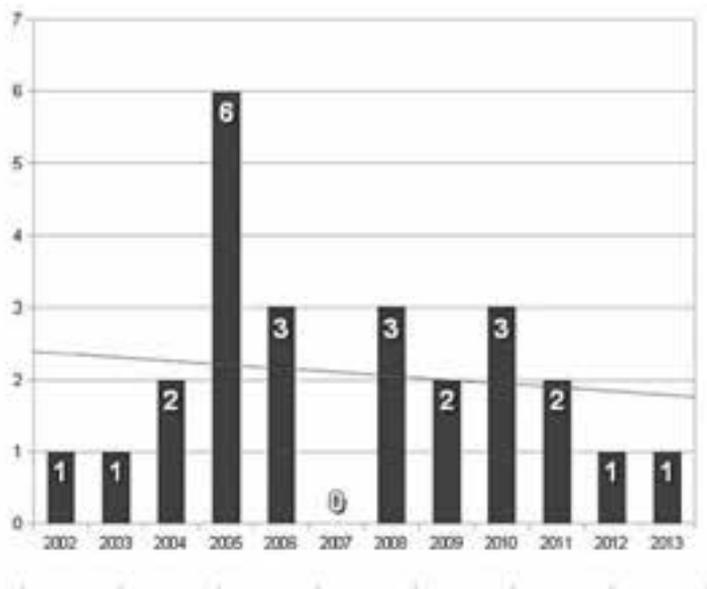
Figure 1. *Registration dynamics of new Wikipedia language sections in the regional languages of the Russian Federation*

---

[136] *CMS – Content Management System – is a computer programme which maintains and organizes joint creation, editing and management of the web-site content.*

[137] *Total number of Wikipedia sections in different languages is 275 as of June 2014. These are independent web-sites in essence, with their own policies and communities.*

330

Simultaneously, other CMS MediaWiki-using wiki projects of outsider producers appeared on the Web. Among them are the hosting Wikia service, Gramps genealogic programme web-site, Arctic Megapedia – an encyclopedia in several minority languages, Olonkho Portal – a web-site dedicated to the Sakha epic, and others.

The history of wiki-projects in the regional languages of Russia goes back to 2003, when the first wiki-project of the kind, the Wikipedia section in the Tatar language[138], was registered. The following year, the Chuvash Wikipedia was registered, and in 2005 there were already six language sections. By 2014, there were 25 sections in the regional languages of Russia (see Figure 1). Today, the list may be extended by Wikipedia in Crimean Tatar which was registered in 2008. In addition, 41 language sections are under trial, which means that these projects do not have the third-level domain yet, but articles for them may be written in a special section called Incubator[139].

In addition to Wikipedia, the regional languages of Russia are represented in some other wiki-projects of Wikimedia Foundation which are produced in 23 languages (see Table 1).

Table 1. Wikimedia Foundation wiki-projects in the regional languages of the Russian Federation (as of June 2014)

Project Name	Registered Projects	Declared Trial Sections
WikiTextbook	2 (Tatar and Chuvash)	2 (Bashkir, Ossetian)
WikiDictionary	1 (Tatar)	15 (Avar, Bashkir, Buryat, Veps, Votic, Ingush, Komi, Moksha, Mari (meadow), Tatar, Tat, Chuvash, Chukchi, Udmurt, Erzya, South-Yukagir)
WikiCitations		3 (Bashkir, Veps, Tatar)
WikiNews		2 (Chuvash, Tuvinian)
WikiGuide		2 (Tatar, Veps)
WikiTeca	1 (Sakha)	12 (Adygei, Bashkir, Veps, Mari (mountain), Kumyk, Mari (meadow), Komi-Permyak, Karelian, Tatar, Chuvash, Udmurt, Erzya)

---

[138] *For reference: Russian Wikipedia appeared in 2001.*

[139] *See the full list in Russian Wikipedia in the article "The Project of Wikipedia Sections in the Regional Languages of Russia/Statistics".*

The wiki-projects under the aegis of Wikimedia Foundation are implemented based on the crowdsourcing model[140] with inviting volunteers on a nonprofit basis. The volunteers do not have to meet any special requirements or qualifications. The formula of success of all wiki-projects is the principle of transformation of quantity into quality; each article (sometimes each work) is published under the open license and can be easily altered by other participants. At that, if we speak about copyrighted works (this concerns mainly the Wikiteca – a library of texts that have entered public domain), non-property rights are preserved in full.

The nonprofit crowdsourcing model supposes that volunteers are motivated by pure nonfinancial reasons. Therefore, if we speak about minority language projects, we can suppose that volunteers are motivated, at least partly, by their love for their mother tongue, and desire to preserve, develop and promote it to the Internet (in some cases, one can speak of language revitalization).

Regional Languages and Drivers of the Wiki-Projects Development

It is possible to estimate all Russia's regional languages according to the degree of their functional development to the beginning of Web 2.0 domination, since this is the milestone which will determine the future of languages. Based on this very criteria, regional languages of Russia may be divided into two big groups:

1. Languages that have managed to accumulate the literary norm and enormous text and media material in different genres (first and foremost, fiction and journalistic literature) to the time of Web 2.0 intervention. At the present-day stage of development, these languages master (to a different degree of success) new spheres including ICT, a metasite. Such languages may be called *successful regional languages*. Among them are the official languages of some republics making part of the Russian Federation.

2. Languages that lack a consistent written and literary tradition or those that have lost their status to a large extent. Though some of them boast of formal achievements, they are functioning at a household level only. This group includes both minority languages and the languages of relatively large ethnic groups (up to 500,000 people), but the amount of fiction produced in these languages is very low and written journalism is almost invisible (though electronic mass media – TV and radio – may be developed relatively well). Such languages may be called recessive regional languages. They include minority languages and, with some allowances, the languages of relatively numerous groups that have official status in the regions of Russia.

---

[140] *Crowdsourcing is a joint volunteering work (which is often coordinated by means of ICT) or a joint work on the web-site content or software localization.*

Wiki-Project Development Scenarios

There are two opposite scenarios for wiki-projects in the regional languages of Russia, as there are for all crowdsourcing Internet-projects in these languages:

Scenario 1. Web-sites in the regional languages of Russia are made in the grammatically correct languages, with full observance of literary norms that were refined in the Soviet period, during the political reforms in Russia at the end of the 20th century and, partly, recently (including the so-called reactionary changes that were typical of some languages in the second half of the 2000s and recent 2010s).

Along with the literary norm, the web-site content will contain, though to a different degree, dialectisms, phonetized internationalisms[141] and occasionalisms. In some projects and texts, journalistic ones in the first place, there may occur traditional phraseological units (for instance, in Wikinews[142]). However, since the usage[143] will more and more deviate from the literary norm (see below), Wikipedia and similar projects in the regional languages (we are not going to dwell upon the popularity of the projects in the dominating languages, for example, about Wikipedia in English) won't be generally popular among the representatives of the respective ethnic groups, though they will be demanded by some enthusiastic patriots, linguists and cultural studies scholars. Occasionally, these projects may be demanded by politicians. Such scenario may be called negatively correct or pessimistically traditional.

Table 2. Two wiki-project development scenarios for minority languages

Scenario	Developer	Forms of realization	Language	Consumer	Mission
Pessimistically traditional	Good proficiency in language, knowledge of culture	Traditional external forms – written	Literary language	Scientists, politicians, students and their parents, journalists, social activists	Preservation of culture and language

---

[141] *Phonetized internationalisms are internationalisms that have been changed and processed phonetically by the language.*

[142] *Wiki News is a news web-site run by volunteers. It is produced in several languages, including Russian.*

[143] *Usage signifies that language units (words, word combinations, forms and structures) are used in the conventional manner by mother tongue-speakers.*

| Populist | Common language, sometimes oral only | Media forms (images and sounds) prevail | Newspeak (unchanged borrowings: syntax structures and vocabulary) | Average consumers, jingos | Formal preservation of ethnic groups in the changing environment |

Scenario 2. This scenario is more optimistic from the standpoint of project success (demand and the number of active and passive users); but, from the standpoint of language (language realization), it is not so optimistic. According to Scenario 2, wiki-projects will follow the changes which are taking place in the languages (moreover, they will serve these changes, thus contributing to their codification). There are grounded doubts that these projects will be written using simplified language versions which remain under strong impact of the dominating language, up to creolization[144], and since creolization of regional languages is typical of a significant active and younger part of ethnic groups, these projects will find a larger target audience and, consequently, will be more popular than Scenario 1 projects. This Scenario may be called conventionally optimistic or optimistically defeatist, or merely populist.

As we can see, though the development scenarios are opposite in the form of execution, it may be hard to give preference to any of them from the standpoint of language vitality because, in case of Scenario 1, the language remains relatively unchanged but, as the time goes by, it loses (or doesn't acquire) potential speakers; in the second case, the language acquires more potential speakers but faces the risk of changing drastically and, most probably, in the direction of simplification and creolization.

Correlation between the Development Scenario and the Current State of the Language

Is a development scenario determined by the current state of the language? The Wiki-projects in the languages of Group 1 (successful regional languages) may develop according to one of the two above mentioned scenarios depending on the decisions of the customers/executives (by saying that we do not mean institutional customers in the ordinary meaning because we are speaking about volunteer projects). The Wiki-projects in the languages of Group 2 (recessive regional languages) may develop in the traditional pessimistic direction. Such

---

[144] *Creole languages are superstrata languages or pigeon languages which have become mother tongues for specific communities.*

development is stipulated by conservation tendencies of the said languages since these they have been scientifically described (or documented) but have lost (or will lose in the near future) a chance for changing as a result of the natural decrease in the number of their speakers. In the social environment, such languages will retain the role of a token showing the adherence to the cultural elite of the ethnic group.

In other words, successful regional languages retain a chance for selecting the development scenario while recessive regional languages are deprived of this chance.

Cutting Direct and Inverse Links between the Literary Language and Usage

In this chapter, we are approaching a phenomenon of significant divergence that exists between the usage norm in the successful regional languages and the literary norm. Why is it so? Remember that usage, on the one hand, is stipulated by the literary language, but, on the other hand, it impacts this literary language. In ordinary conditions, these two norms coexist as communicating vessels: they complement each other and interact.

In our opinion, these processes tend to weaken. The literary norm of a minority language stops being demanded as a result of cancelling the teaching of this language as a mandatory one at secondary school (or even complete cancellation of its teaching). Thus, the representatives of the ethnic group lose the habit of reading and writing in their mother tongue. Though the literary language may still be preserved on some "islands" (TV and radio, for instance), where it can be perceived aurally, this status-quo will be preserved only *in case of a political will to preserve these quasi-witnesses of the political independence* of the regional elites. As a result, the number of radio and TV programmes in different minority languages also tends to decrease because *ethnic units* master the dominating language(s); because they don't know their mother tongue well enough; and because the majority of the ethnic group deviate from their traditional spheres of economic activity (e.g., agriculture) where the particular language was historically self-sufficient.

Note: even these islands of wellbeing witness gradual penetration of the newspeak. At first, the reason for it is the intention to attract young audiences; then it becomes virtually impossible to follow the literary norm – titanic efforts are required for that because fewer and fewer people on the staff speak these languages. However, this penetration won't impact the formation of the new literary language because of small audiences and absence of codified changes.

In such conditions, there comes a logically inevitable moment when the literary norm stops refining the usage and keeping it within the formal framework. In new conditions this unwritten code stops working to a significant extent because the decreasing number of readers leads to the "migration" of writers to the dominating languages, and the literary norm stops being regenerated because of the absence of new literary writings. The literary language will change insignificantly but the codification of the new words and the development of this process may be hindered by linguists, no matter how paradoxically it may sound. By this time, linguists will have remained virtually the only users of the literary norms (writers and journalists will no longer exist because of the absence of demand in them). Linguists are professionionally puristic[145]. They are not open to changes, because, by definition, they are entitled to describe the processes and not to have an impact on them. Moreover, it will be economically unprofitable for them to contribute to the changes in usage because they will remain rare and highly estimated experts.

Nevertheless, it's worth noting that, in real life, there will exist a mix of two scenarios. These scenarios might be realized for different languages according to the earlier made plans provided there are highly motivated communities capable of organizing this langiage preservation movement, but, in all likelihood, these scenarios will be realized spontaneously.

Conclusion

Starting to develop new wiki-projects in regional languages, volunteers should bear in mind the tendency towards unilateral migration of regional languages in the direction of dominating languages. There are two different approaches to the development of crowdsourcing linguistic projects. The choice depends on both the state of the language and the goals of the project/team. Perhaps, in order to save the resources, the team should adhere to one of the two opposite approaches or try to combine them and choose the mediana way by not allowing any of the approaches to win. The understanding of this paradigm is essential for setting real goals and tasks, improving the performance and preventing conflicts both inside and outside the project. One thing is obvious: there is no way to ignore the development matrix which is stipulated by impartial factors.

---

[145] *Purist is one preoccupied with the rigidity and purity of a language.*

Nestor RUIZ

Lecturer and Researcher, Caro and Cuervo Institute
(Bogota, Colombia)

Raising Awareness in Cyberspace about Colombia's Linguistic Diversity:
The Experience of the Instituto Caro y Cuervo

1. Introduction

According to the surveys for the Sociolinguistic Atlas of Latin America Indigenous Peoples (2010), Colombia is the second most linguistically diverse country of the American continent, between Brazil (country that leads the list) and Mexico (that follows). That diversity is represented in 62 indigenous languages spoken currently in the national territory, along with two creole languages (one is Spanish-based, while the other is English-based), two varieties of Romani, and Spanish as the language of the majority of the population, recognized as the only official language of the country and present in all aspects of the society, from the media to the education system.

2. Colombia's Linguistic Diversity

2.1. Indigenous Languages

Colombia's linguistic diversity was first recognized by the Spanish Conquistadors, who encountered a territory that (to their eyes) seemed like a re-erected Babel. As the formal conquest of Colombia's inland territory started, around 1530 A.D., the Conquistadors were well aware of the existence and use of Nahuatl, Quiche and Quechua (the major and most spread languages of America at the time of the contact with Europeans); numerous priests, captains and soldiers had knowledge of those languages and served as interpreters on the expeditions and military campaigns. But, given that the territories of Colombia were far outside the Inca and Aztec or Maya rule, and that –as some historians have pointed out– the country was for centuries a place of contact and influence of four different American cultures (Amazonic, Andean, Caribbean and Mesoamerican), what the Conquistadors found was an intricate and overwhelming puzzle of languages and ethnic groups, where no known language was to be found or interpreter

be of any use. One of the most cited impressions of this diversity is that of Pedro Cieza de León, a Spanish chronicler who traveled the west portion of the country from 1538 to 1548 and wrote "*Hay tanta multitud de lenguas entre ellos que casi a cada legua y en cada parte hay nuevas lenguas*" [They have so many languages that with almost every league walked or village encountered comes a new language][146].

Some colonial-times historians considered that more than 140 languages were spoken in the actual Colombia around 1530. Today, and based on linguistic research, it is estimated that at that time, over 80 languages with their own dialectal varieties were spoken in the territory, grouped in 13 linguistic families[147] (Cfr. Ortiz 1965: 395). Of those 80 or more languages, 62 survive today in various degrees of vitality and use. Those surviving reached the 21st century thanks to the isolation (relative in some cases, extreme in others) of their populations. The four strongest and widespread indigenous languages spoken during the conquest (*Muisca*, *Tayrona*, *Pijao* and *Andaquí*) died over the colonial period and did not reach the 19th century, mostly due to the Demographic Catastrophe of the conquest and later to the enslaving and acculturation of Indians.

Of the 62 surviving languages, 28 are classified by the SIL, UNESCO, and many other institutions, as "Endangered", "Threatened", or "Dying", and their populations always count less than a thousand, or even a hundred people. But, the major threat to the strengthening and diffusion of these languages, either endangered or not, is the consideration (shared nearly by all indigenous populations) that they do not offer any opportunity for social or economic advance, contrasting with Spanish, which is seen as a prestige language and a key feature for having a chance of success in the today's world; furthermore, the fact that the entire educational system of Colombia is Spanish-based presses deeper this issue.

Below we offer a table that lists the indigenous languages, their linguistic filiation, and current state[148].

---

[146] *Translation is mine.*

[147] *Many of the first considered by the Spanish as "languages" were later classified as varieties, and ethnographic and archeological studies have shown that there were certain (and in some cases strong) ties between ethnical and linguistic groups.*

[148] *The linguistic filiation was adapted from González de Pérez [2011]; the evaluation of the state of the language from Ethnologue.com.*

Table 1. Classification and current state of Colombia's indigenous languages

Linguistic Family	Language Name	Status
Arawak	Achagua	Threatened
	Baniva	Threatened
	Kabiyarí	Shifting
	Kurripako	Developing
	Piapoko	Developing
	Wayuunaiki	Developing
	Yukuna	Developing
Bora	Bora	Shifting
	Miraña	Shifting
	Muinane	Shifting
Chibcha	Barí	Vigorous
	Chimila	Threatened
	Damana	Threatened
	Arhuaco	Developing
	Kogui	Vigorous
	Kuna	Developing
	Tunebo	Nearly Extinct
Choco	Embera	Developing
	Waunan	Developing
Guahibo	Guayabero	Developing
	Hitnu	Threatened
	Cuiba	Developing
	Guahibo	Threatened
Karib	Carijona	Nearly Extinct
	Yukpa	Threatened
Quechua	Inga	Threatened

Macú-Puinave	Puinave		Threatened
	Hupdë		Threatened
	Kakua		Developing
	Nukak		Vigorous
	Yuhup		Unknown
Peba-Yagua	Yagua		Unknown
Sáliba-Piaroa	Piaroa		Shifting
	Sáliba		Shifting
Tukano	West	Koreguaje	Threatened
		Siona	Shifting
	East	Waimaha	Developing
		Barasana	Developing
		Desano	Threatened
		Carapana	Developing
		Cubeo	Vigorous
		Makuna	Threatened
		Piratapuyo	Threatened
		Pisamira	Threatened
		Siriano	Threatened
		Tanimuca	Threatened
		Tatuyo	Developing
		Tuyuca	Threatened
		Wanano	Threatened
		Yurutí	Developing
Tupí	Cocama		Nearly Extinct
	Yeral		Shifting
Uitoto	Ocaina		Moribund
	Uitoto		Developing

	Andoque	Shifting
Not classified / Unrelated	Awa-Cuaiquer	Threatened
	Guambiano	Developing
	Camsá	Developing
	Cofán	Threatened
	Nasa-Yuwe	Threatened
	Ticuna	Threatened
	Tinigua	Nearly Extinct
	Totoró	Dormant

People speaking these languages, a number around 1,600,000, account for 3.4% of the country's total population. Four indigenous communities, and their languages, are currently the largest and strongest: the *Wayuu* people in the north coast (speakers of *Wayuunaiki*); the *Nasa* people (speakers of *Nasa-Yuwe*) and the *Guambiano* people (speakers of *Guambiano*), both on the highlands of the Andean southwest; and the *Arhuaco* people (speakers of *Arhuaco*) on the coastal range of the Sierra Nevada de Santa Marta, in the north coast. All of them have a common feature that explains their strength: they have created very resilient political projects and resistance discourses which have laid out the basis for their consolidation.

2.2. Creole Languages

Two creole languages are spoken today on the territory of Colombia: *Palenquero*, which is a Spanish-based creole, and *Criollo sanandresano*, which is an English-based one. The *Palenquero* is the only Spanish-based creole currently in use on the American continent, while the *Criollo sanandresano* is part of the many English-based creoles scattered through the Caribbean islands (the islands of *San Andrés* and *Providencia*, where this creole is spoken, came to be part of Colombia's territories after the Independence Wars (1809–1819), before that, they were part of Nicaragua's territories, though they never were under its rule)[149].

Palenquero is spoken today by a group of around 3500 people, located in a tiny village of San Basilio de Palenque, some 50 kilometers southeast of Cartagena de Indias, on the north coast of the country. The village was formed as a *Palenque* (name that was given to the slave villages outside of Spanish rule,

---

[149] *These islands became famous since colonial times, as they were, from 1670 to 1680, the base for two infamous pirates: Henry Morgan and Edward Mansvelt, who incorporated the islands de facto to the English colonies of the Caribbean.*

a fact that also explains the name of the language) during the 17th and 18th centuries; it was erected mainly by maroons, but some historical studies have shown that among the regular population there were also Spanish and mestizo looters, native Indians, and even Spanish and mestizo women accused of witchcraft. The resulting community created a creole language lexically based on Spanish but with features of African languages of the maroons, who were the vast majority of *Palenque's* people.

Today many *Palenqueros* speak the Creole as their mother language, but the percentage of bilinguals in Creole and Spanish has risen up to 93% of the population in the past decades; Spanish is a hard competitor for *Palenquero*, and most of the sociolinguistic surveys conducted over the past 20 years have concluded that the Creole is in a path towards weakening and ultimately extinction (see for example, Patiño Roselli 1992). Ethnologue.com classifies it as a "Moribund" language, and we can go further and point out that the Creole is living its last days in a situation of "Diglossia": both Spanish and Creole have their own and different rules of use, and the Creole is reserved only for intimate interaction, while Spanish is used as the main vehicle for communication. Although the older generations of *Palenque* feel very proud of their tradition and speak the Creole as their main vehicle of communication, the younger generations are thinking in a different direction and quickly abandoning the *Palenquero* in favor of Spanish, the "Ethnic Shame" being the main reason for this decision, along with the consideration that Spanish is a key tool for social and economic advance, a surplus that *Palenquero* lacks.

On the other hand, the *Criollo sanadresano*, as happens with the Caribbean creoles, is English-based, and seems to be closely related to the *Jamaican Creole,* the *Belize Kriol* and the *Miskito Coastal* creole. The most-cited classification features it as part of the Western Caribbean creoles. It is currently spoken by 20,000 to 25,000 people who, as it was said before, live in the islands of *San Andrés* y *Providencia* forming a strong and very robust ethnic group; the inhabitants born and raised on the islands call themselves *Raizales* (something like "*Root people"*) and are really proud of their ethnic, linguistic and cultural heritage.

Almost all native islanders are trilingual and fluent in Creole, standard English and standard Spanish. While Spanish and English are the languages for commerce, institutional affairs, and tourist services, the Creole is reserved for familiar and intimate communication; nonetheless, it plays a key role in keeping strong social ties among the islanders. Another fact that has contributed to the maintenance of the *Criollo Sanandresano* is that most of the islanders belong to the Baptist church, which offers religious services in English and Creole, thus creating a link between the use of the language and the spiritual life of the population.

The linguistic situation of this creole is very different from *Palenquero*, and resembles more a "Creole Continuum" in which a speaker is able to move from a "Basilect" (the creole language itself) and an "Acrolect" (standard English for the case), selecting the variety that better fits the communicative situation or the interlocutors. Many islanders, though, don't recognize their language as a creole, but rather prefer to think of it as a "very fast utterance" of standard English. It is no surprise then that the status of the Creole, according to Ethnologue.com, is "Vigorous".

2.3. Romani Language

In Colombia there is a small but strong community of Romani people (also known as Gypsies or Gitanos) who use and maintain their ancestors' language. In fact, there is not one, but two groups of Romani in Colombia: the largest, of proper Romani people, and one smaller, of *Ludar* (or Romani who are Rumanian-related) people. The strongest clusters of Romani people are located in the cities of Cucuta (in the border with Venezuela), and Pasto (near the border with Ecuador), where they dedicate, mostly, to international (and somewhat informal) trade. T   here is another important cluster in Bogotá D.C., the capital city, and in many other second- or third-level cities of the country. They are a number around 5,000 to 8,000 people, and many of them face the hostility of the majority of the country's population, who considers them as thieves, counterfeiters, smugglers and even drug dealers. Some of the small Romani communities also experiment "Ethnic Shame" and never use their language in public or don't even admit that they are fluent speakers of it.

Surprisingly, the linguistic situation of Colombia's Romani language is that of "Bilingualism". Caballero (2001) points out that the language of familiar interaction can be either Romani or Spanish with a slight preference for Romani. Children learn both Spanish and Romani in their houses, and do not perceive Spanish as a prestige language or a tool for social or economic advance. Backing up of those facts, Ethnologue.com classifies the Colombian variety of Romani as "Vigorous", and some sociolinguistic research have concluded that, given the strong identity of the group and the long (an also vast) cultural heritage they share, the Romani language faces a struggle with Spanish, but is not threatened or on its way towards extinction.

According to Caballero (cit.), none of the Romani people of Colombia is monolingual in Romani, but bilingual in Spanish and Romani. That bilingualism offers a striking fact: most of the Colombian Romani have two first names: one in Romani and other one in Spanish, functioning like a "Real" (Romani) name and an "Institutional" (Spanish) name. In most of the cases either name is

completely unrelated to the other. In a risky approach to the situation, one could say that the bilingual status of the Romani speech community in Colombia has granted the survival of the Romani and its passing to future generations.

3. Colombia's Linguistic Diversity in Cyberspace

3.1. *The Instituto Caro y Cuervo's Initiatives*

The Instituto Caro y Cuervo (ICC) is a Colombian research and education centre focused exclusively on studies of language and literature; it is recognized in the Spanish-speaking (and academic) world as one of the top research centres in America, and holds two of the most prized recognitions for academic excellence in Ibero-America: the *Príncipe de Asturias* award and the *Bartolomé de Las Casas* award. Founded in 1945, the ICC owns its fame and recognition for its contributions to the studies of Spanish language, and during its first fifty years of existence that was its main (if not only) focus of research. However, even from the first years the ICC addressed the recognition of the linguistic diversity of the country, and manifested interest in the study of all the languages spoken in Colombia, along with the problems that raised from language contact, a tendency that grew strong over the years.

As new times come with new challenges, the technological revolution of the 20th century led the ICC to gain interest in applying digital tools for the research in social sciences, especially in linguistics and literature, and ultimately, that interest crystallized in recent years as a new social mission for the institution: to gather and maintain the linguistic patrimony of Colombia, taking advantage of the newest electronic tools and spaces, such as the Internet. And it is, precisely, in that quest for gathering and maintaining the linguistic patrimony of Colombia that emerges on late 2008 a project for an Internet portal, hosted at the ICC servers, intended to display the linguistic diversity of the country and serve as a tool for linguistic planning, among other objectives.

The project *Portal de lenguas de Colombia* ("Colombia's languages portal") was born in early 2009 and has been in continuous service (and growing strong) since then[150]. The portal is coordinated by Yaty Andrea Urquijo[151], a researcher for the ICC specialized in language diversity and language contact. The portal was conceived as an effort to raise awareness about Colombia's linguistic diversity while getting advantage of the new tools given by the Internet and other new technologies. A very complete, thorough

---

[150] *Available at http://www.lenguasdecolombia.gov.co/.*

[151] *yaty.urquijo@caroycuervo.gov.co, yatyurquijo@gmail.com.*

and growing website depicts in detail the country's linguistic situation and gives information on many levels about the languages currently spoken. All the information, at the time, is available only in Spanish.

The portal offers many contents, all of them revised and checked from the linguistic perspective to assure its accuracy. It has four main features: the first one is the *Map of Colombian Languages*, which is a cartographic interface that geographically localize any given language over the national territory; once the user selects one language an information window emerges that shows all the pertinent information for it: territory where it is spoken, linguistic filiation, population and number of speakers, a short review on the state of the language, a short ethnographic survey, and, if available at the time, some audio clips and pictures to illustrate ethnographic or linguistic aspects. Besides the information provided by the map, every language included in the portal has also a detailed file which includes access to scientific articles, links to other references, and further orientation for the interested person.

The second feature is the *Logbook*, a blog dedicated to publishing field notes and on-site observations (and even raw linguistic or ethnographic materials) taken by investigators of the ICC or other institutions on their fieldwork with Colombian languages. It offers a personal and sometimes intimate view of both the community (or language) studied and the daily-life experience of the investigator.

The third and perhaps most ambitious feature of the portal is *The Interactive ALEC*, ALEC being the short name for the *Atlas Lingüístico-Etnográfico de Colombia* (Colombia's Linguistic and Ethnographic Atlas), a massive six-volume linguistic atlas focused on the geographical variation of Colombian Spanish, considered a true milestone in Spanish linguistics, that was for thirty years (from its publication in 1983) the only completed linguistic atlas for a Spanish-speaking country. The ALEC has 1532 single maps bursting with linguistic and ethnographic data, and the total number of lexical items that can be consulted on those maps can easily be over the million, or million and a half (in fact, there is no official count of all the lexical elements present on the ALEC).

With such enormous numbers in mind, *The Interactive ALEC* is basically an effort to convert to a digital format all the data available in the six volumes of the original work, a task that has proven to be intricate and highly challenging. The idea is that the ALEC can be available and accessed through the Internet for everyone interested, but the daunting number of lexical elements, along with geographical and cartographic questions about the management of the data, has posed a slow rhythm to the development of

the feature, and sometimes forced to stop all the work, rethink the situation in light of a new problem discovered (or a solution available) and start again from zero. Because of that, only around 15% of the original material has been uploaded, formatted and made available through the portal. Currently the team is working hard not only to make possible the electronic visualization of the cartographic data of the ALEC, but to advance the development of tools that allow to work with the data in terms of specialized searches, cross-referencing of linguistic and geographical variables, or construction of a system of annotations for collocate searches.

The fourth and final feature of the portal is the electronic journal "*Lenguas en Contacto*", which is a non-periodical journal published on the site, dealing with the problems, characteristics and outcomes of language contacts and linguistic diversity in Colombia. Five numbers have been published so far with the participation of investigators from the ICC and other Colombian institutions. Those interested in these subjects can find there full articles, reviews, notes and materials of value, though it is all available in Spanish at the time.

4. Discussion

With all the features discussed, the portal, then, summarizes, describes and analyzes Colombia's linguistic diversity, and serves as a showcase for the country's languages and their mutual relationships. It covers, as we have been doing on this presentation, the indigenous languages, the Romani languages, the Creole languages and the Spanish language and its 13 geographic variants. According to the team coordinator, the portal receives more than 300 visits every month, a number that can seem modest at first but becomes significant if we consider that the contents published, the objective public and the general orientation of the project are aimed at a very specific and specialized profile, that of a linguist or an anthropologist interested in the study of language.

In that same sense, it has to be noted that most of the traffic that the pages receives comes from academic institutions, investigators, professors and students of language, linguistics and anthropology. That "specialization" of the portal, both on the side of the contents and on the side of the visitors, tell us a great truth: the *Portal de Lenguas de Colombia* is more a tool for academic and scientific research and reference, than a consistent effort for increasing the knowledge of the general public about Colombia's linguistic diversity. The team in charge of the project came to the same realization after two years of work, and decided to start searching for new approaches that could capture the attention of the general public and raise awareness about the country's languages and ethnic groups.

One of the first and best results of this new approach is the recent "Kid's Section" of the portal, whose content is less specialized and more oriented to the ethnographic and communicative aspects of the languages. This section also has a very attractive graphic design, specially developed for kids, where plenty of drawings of typical dresses, houses, and customs (among others) of the ethnic groups illustrate the linguistic and cultural diversity in a very colorful manner. Every ethnic group is represented by a kid who bears the distinctive marks of his people, starting with his first name. The idea is to illustrate how those kids (and hence those ethnic groups) are able to interact, work together and solve the most common problems of a country with a high linguistic diversity (which are, namely, the underrating of non-dominant languages and ethnic groups, and the resulting glottophagia and even ethnocide in favour of the language and the groups that bear the main recognition).

The team is projecting to enrich the Kid's portal with games, music and pictures, so that visitors can feel more involved in (and understand more thoroughly) the subject of the linguistic diversity of the country. So far, this has been the most consistent (and fruitful) path to turn the *Portal de Lenguas de Colombia* in a tool for promoting, recognizing and sharing the facts about the linguistic diversity of the country. Paradoxically, many of the first visits for the Kid's Portal came from basic and middle school teachers who assigned to their students reports on Indigenous, Creole and Romani languages, indicating that the information was provided by the Portal.

Consequently with all the above, the team has also encountered a problem of visibility of the *Portal*: its existence and contents are well known among language teachers, researchers, and related institutions, but not with the general public, who has no idea that such a tool exists and is available for consulting. Another fact plays a crucial role here: that the Instituto Caro y Cuervo is generally perceived by the public as an "Ivory Tower" of Spanish studies, detached from the country's reality and more interested in monumental and highly specialized studies, or in the purism and correction of the Castilian language, than in the preservation of national culture and diversity (a misconception that we already mentioned above). In that sense, many people find it hard to believe that the ICC wants to preserve and document the linguistic patrimony of Colombia, and that the same institution hosts a web portal that showcases in detail the country's linguistic diversity. The last sentence expresses another big problem of the portal: it is a tool for showcasing, but not for exploitation of the country's linguistic diversity. So, it is our challenge to turn the *Portal de Lenguas de Colombia* from a showcase tool for linguistic diversity into a productive initiative aimed to take advantage of the country's linguistic diversity on its favor.

As a result from the presentation of this paper during the 3rd Yakutsk Conference, three very important observations were made by the attendees: 1. The Portal should be available in English so that a wider audience can be targeted. 2. The Portal coordinators should start to think of the speakers of Indigenous, Romani and Creole languages as collaborators, not mere informants or providers of linguistic or ethnographic data, so that a real interaction can be established between the communities and the academia that sponsor this project. 3. The Portal should evolve in the future into a cybercommunity, incorporating tools such as forums, chats, Q&A, and even live streamings, so that the interested public can really interact and share knowledge, rather than simply access it through the site. We at Instituto Caro y Cuervo believe that these suggestions are really valuable and pertinent not only to our experience, but to other similar projects currently ongoing in the world.

As UNESCO, the UN and other institutions have come to realize the linguistic and cultural diversity of a country is an asset that can be used in favor of the creation of better social conditions; the IFAP Programme points out that the linguistic diversity of a country can be directed towards the construction of a more tolerant world. In that sense, and as a final remark, we would like to ask to all the fellows who participated in this 3rd International Conference, the UNESCO, and the IFAP Programme: given your own experience, how can you, your institutions or your experience, help us (and other similar projects) to face this challenge?

References

1. Caballero Rodriguez, O. (2001). Aproximación sociolingüística a la comunidad Gitana Rom de Colombia. In: *Revista Forma y Función,* N° 14., pp. 67–82. Universidad Nacional de Colombia.

2. Ethnologue, Languages of the World. Available at: hhttp://www. ethnologue.com.

3. Gonzalez De Perez, M. E. (2011). *Manual de divulgación de las lenguas indígenas colombianas.* Bogotá, Instituto Caro y Cuervo.

4. Ortíz, S. E. (1965). Lenguas y dialectos indígenas de Colombia. In: Martínez Delgado, L. (Coord.). *Historia extensa de Colombia,* t. III, v. 1. Bogotá, Ediciones Lerner.

5. Patiño Roselli, C. (1992). La criollística y las lenguas criollas de Colombia. In: *Thesaurus, Boletín del Instituto Caro y Cuervo*, t. XLVII, N° 2, pp. 233–264. Bogotá, Instituto Caro y Cuervo.

6. Schira, I. (Coord.) (2010). *Atlas sociolingüístico de pueblos indígenas de América Latina.* Cochabamba, FUNPROEIB Andes / UNICEF.

the collapse of the Soviet Union, the status of the Kazakh language raised the all-national spirit. It is remarkable that the domestic Kazakh speaking environment used to be limited only to the family, but now it is developing in all spheres of life.

The Kazakh language is a national phenomenon and is full of vital energy. Nevertheless, we believe that the national intelligentsia has to consolidate its efforts in the development of the mother tongue. There is a number of scientific and applied problems in this sphere.

First of all, they are connected with the terminology of the Kazakh language, as there is no common approach to the rules and usage of various terms. We also face the problem of the adequate translation into Kazakh. It is necessary to adopt the law on the state language. The state program of the development and functioning of languages in Kazakhstan for 2011-2020 is aimed at the solution of these problems.

The newly adopted conception of the state language policy has become a logical continuation of the consistent and step-by-step support of the state language in the country. We believe that the Kazakh language is going to get a status of a language of international communication. It already functions in all spheres of our life.

In the era of globalization, it is difficult to underestimate the importance of the knowledge of several foreign languages. The knowledge of foreign languages is not only a basic element of the integration into the world community, but also a factor of the personal competitiveness.

It is remarkable that the First President of the Republic of Kazakhstan Nursultan Nazarbayev proposed the idea of the trinity of languages. He believes that the knowledge of the Kazakh, Russian and English languages is a guarantee of the successful future of every Kazakhstani citizen. Back in 2004, the President mentioned that it was necessary to learn the English language. In 2006, at the XII session of the Assembly of the People of Kazakhstan, the President spoke about the cultural project "Trinity of Languages". In 2007, in his message to the people of Kazakhstan "New Kazakhstan in the New World", the Head of the State proposed a step-by-step implementation of the "Trinity of Languages" cultural project. Since that time, the trinity of languages has become the priority of the state language policy.

declaring independence, all nationalities of Kazakhstan settled down to a course of economic, social and cultural development and have been pursuing this course together ever since. As was mentioned before, Kazakhstan's population is multinational. Each nation has its history, traditions, culture and religion. Among permanent residents of Kazakhstan are representatives of *over 100 ethnic groups*. The most numerous of them (in descending order of population) are Russians, Uzbeks, Ukrainians, Uyghurs, Tatars and Germans.

On March 1, 1995, The Assembly of the Peoples of Kazakhstan was established. It is a unique institution of tolerance that unites numerous ethnic groups of the country and has no counterparts in the Former Soviet Union. Peace, concordance, safety, solidarity and stability are the key values to be cherished and preserved by the Assembly.

However, in spite of the multiethnic nature of the country, Kazakhs are the title nation of Kazakhstan, the nation which is related to its national development. Since the development of the language policies remains under the impact of globalization, Kazakhstan is facing the task of preserving its national culture and language without allowing them to dissolve in the global culture of better economically developed countries. Sociologists claim that globalization leads to minimization of the role and significance of national states which will be replaced by supranational organizations. However, judging by the facts of life, nations and national states can't be removed from the historical arena. Therefore, globalization has to be considered at the level of inter-ethnic relations. The countries of Middle Asia, which declared their independence after the decay of the USSR, and the European countries, such as Croatia, Slovenia and Macedonia, are a visual proof of this statement. Realizing all the dangerous tendencies of globalization for the national culture, we can't opt out of it completely. At the meeting with the editors-in-chief of some Kazakh publications, President Nazarbaev said that we must think of not how to oppose globalization but of how to adapt to it. It is only in this way that we'll be able to preserve our language, religion, traditions and customs.

At the intersection of global civilization and national culture the purpose and task of any thinking nation is to preserve its mother tongue as the core attribute of culture. The language is the core of the Kazakh culture, which makes part of the Turkic civilization. The language is the basis of parenting, the source of knowledge and science, the soul of the nation. Our poet Kadyr Mirzaliev said that every nation, like every human, has its soul. The soul and language are inherently interconnected. When deprived of the mother tongue, the nation will inevitably lose its national flavor. Thus, the unique character of any nation is enclosed primarily in its mother tongue.

Nevertheless, in 2007, the government of the country adopted the Language Trinity project for three languages: Kazakh, Russian and English. The President's Address reads that Kazakhstan must come out on the international arena as a country whose population is well educated and proficient in three languages: Kazakh, the official language of the country, Russian, the language of international communication, and English, the language of successful integration into the global economy. However, mastering three languages is a complicated psychological process. Most of the Kazakh intelligentsia representatives look critically at this project and think it will be an obstacle for the development of the national language. In his latest address, President Nazarbaev outlined the approaches to resolving this problem. When asked if there was a danger for the Kazakh language to degrade under the impact of two international languages, the President said that Kazakh had to become not one of the three but the first of the three languages, the major, dominating and the most important language, the national language of the Republic of Kazakhstan, and enjoy the corresponding attitude and funding. Today, Kazakhstan allocates a lot to the development of the national language. It demonstrates that the government cares of the national language of the country. According to the President, care of the language has to start with care of peace and safety. The President believes that his duty as head of the country and a Kazakh is to contribute to the flourishing of the Kazakh language.

An effective way of developing the Kazakh language is to establish a public fund for its development in compliance with the President's initiative. The administrative support is essential if there is a task of revitalizing language training centres. Based on the Language Trinity project, some primary schools have already started teaching English. However, a child must be educated and must learn the world in his/her mother tongue. Today, we have many "semicultural" Kazakhs *who can speak a little in their mother tongue but aren't either good spellers or speakers*. It is at this point that we start doubting the correctness of establishing the early English teaching programme which may confuse children and shape the so-called new Kazakhs who cannot speak properly in either of the three languages. We have to give a scientific substantiation of teaching languages to children depending on their age. If children are not brought up in the mother tongue environment, they won't serve their people. In order to bring up good citizens and patriots we have to teach children in their mother tongue up to the age of 12. The President said that teaching Kazakh must start at pre-school institutions. A child must take in intellect and knowledge with mother's milk.

As for the Russian language within the Language Trinity project, I can say that Kazakhs are proficient in Russian and this proficiency makes them

informed. Still vital are the words of the great Kazakh poet and enlightener Abai Kunanbaev, "Learn reading and writing in Russian. Spiritual wealth, knowledge, art and other endless secrets are enclosed in the Russian language. To adopt their achievements we have to learn their language and master their science because Russians have become what they are by learning other languages and adopting global culture; the Russian language will open our eyes to the world."

Mastering English as an international language is a duty of every person who strives to keep abreast with the time and remain competitive.

As for the role of the Kazakh language in the Kazakh society, it is clearly stated in the Law "On the Languages in the Republic of Kazakhstan" which reads that the duty of every citizen of the Republic is to reach proficiency in the national language which is the major consolidating factor for the peoples of Kazakhstan.

To conclude, I'd like to note that today, under conditions of globalization, every nation may be capable of contributing to the development of civilization only in case it manages to preserve and develop its national culture, its national identity and its mother tongue.

At the same time, in order to develop knowledge and science within the ongoing integration processes, it is essential to know the major languages of the world, Russian, English and others. As for Kazakhstan, proficiency in Russian is especially significant because Russian has been recognized as the language of interethnic development.

Valentina SAMSONOVA

Director,
National Library of the Republic of Sakha (Yakutia)
(Yakutsk, Russian Federation)

Interregional Information Centre for the Documentary Cultural Heritage of the Peoples of the Russian North, Siberia and the Far East: Contributing to the Preservation and Development of Linguistic and Cultural Diversity

The concept of sustainable development, which was devised for the indigenous minorities of the Russian North, Siberia and the Far East (as of 04 February 2009, No. 132-r), envisages support to be provided to the establishment of multifunctional cultural and enlightening centres, modernization of cultural institutions and creation of information basis for the cultural heritage institutions of the minority peoples.

In accordance with this concept, on November 18, 2010, in Yakutsk, the Federation Council Committee for the North and Indigenous Peoples held a field meeting under the title "Modern Information and Educational Technologies for the Development of Languages, Culture and Spirituality of the Peoples of the North (by the Example of the Republic of Sakha (Yakutia)". The resolution adopted by this meeting reads that an interregional information centre for the documentary cultural heritage of the indigenous minorities of the Russian North, Siberia and the Far East should be established under the National Library of the Republic of Sakha (Yakutia). This centre is supposed to open its branches in the regional libraries of the constituent entities of the Russian Federation where minority peoples reside. Such a centre has to coordinate the joint activity of libraries, institutes, and nongovernmental associations of the northern regions of Russia responsible for collecting, preserving and using the documentary memory of the indigenous minorities of the Russian North.

Following this decision, the National Library did much organizational and preparatory work. One of its achievements was the Interregional Conference and Workshop on "Documentary Cultural Heritage of the Indigenous Minorities of the Russian North: Challenges of Preservation and Accessibility" which were held in Yakutsk in October 2012. Among the participants to these events were representatives of nine regions of the Russian Federation: Nenets Autonomous Okrug, Buryat Republic, Sakhalin Oblast, Krasnoyarsk Krai

and Khabarovsk Krai, Taimyr Autonomous Okrug, Irkutsk Oblast, Magadan Oblast and Moscow.

The participants discussed the issues of collecting, preserving and accessing the documentary cultural heritage of the indigenous minorities of the Russian North as well as preserving and developing the languages of these peoples in the digital environment. Special attention was paid to the role of libraries in the preservation of traditional forms of the self-expression of these peoples. The Conference concluded that this process was curbed by the absence of specific fonts in computer operating systems which enable entering texts and uploading and downloading information in the native languages of the indigenous minorities of the Russian North. It was noted that federal and regional authorities should consider (a) putting publications in these languages on the list of socially significant publications that are produced at the federal and regional expense with subsequent compensation-free handover of a certain part of them to public and school libraries; (b) establishing a special fund to render support to the web-sites and portals in the native languages; (c) contributing to networking cooperation between the cultural and educational institutions and nongovernmental organizations of the indigenous minorities of the Russian North, Siberia and the Far East, and (d) introducing a new subject in the informatics teaching curriculum, namely, the basics of using the languages of the peoples of Russia.

The Conference discussed a plan of implementing the Interregional Centre project and the methods and ways of collaboration between its participants. The workshop, which was held within the Conference, provided training to the participating libraries and taught them to collaborate for the purpose of creating joint information resources. In 2013, several trainings on the methods of library/region collaboration were held in the towns of Ulan-Ude and Yuzhno-Sakhalinsk.

Today, it is possible to say that the Interregional Centre under the National Library of the Republic of Sakha (Yakutia) has set to work. Cooperation agreements have been signed with 21 libraries.

Among the Centre's main lines of action was creating in 2013 the first Russian Interregional Union Catalogue in the indigenous minority languages of the Russian North, Siberia and the Far East. As of today, this catalogue contains 1,692 entries in 24 languages; the number of participating libraries is 21. Replenishment of the Union Catalogue is achieved by means of entries borrowed from the Russian National Library, copied from the catalogues of the National Library of the Republic of Sakha (Yakutia) and participating regional libraries of the Russian North. As of today, 10 regional libraries have created

247 entries for the Union Catalogue. This resource grants access to the fulltext documents of the *Knigakan* e-library on the basis of duly signed publishing agreements.

Knigakan has to become the second corporate information resource created by the partner libraries within the Interregional Centre. Today, this e-library contains 807 documents in 12 languages of the indigenous minorities of the Russian North.

The Interregional Union Catalogue and *Knigakan* are represented on the Knigakan multilanguage web-site which was created in 2010 by the National Library of the Republic of Sakha (Yakutia). In addition, the web-site displays Mediateca, a hall of periodicals with digital versions of the *Tatkachiruk* and *Ilkan* periodicals, subject-oriented collections "Literature Map", "Nomadic School", "Calendar Holidays of the Peoples of the North", "Traditional Households", etc. This web-site is in rather high demand: it has been visited by 18,000 users from the Netherlands, Kazakhstan, the United States, Great Britain, Kenia, Ukraine, Belarus, Germany and other countries.

However, some challenges the Centre is facing are worth mentioning. First, the participating libraries do not perform direct online cataloguing in the Union Catalogue, which impedes the development of this corporate resource. Second, many participating libraries lack material and technical resources and qualified staff to ensure quality digitization of documents. Therefore, we are now setting the task of establishing a Uniform Digitization Centre on the basis of the National Library of the Republic of Sakha (Yakutia).

Hereinabove, we have described only two main directions in the activity of the Interregional Centre which carries out a diverse work aimed at preserving and disseminating the documentary heritage of the indigenous minorities of the Russian North. For instance, the Centre participates in the implementation of the National Bibliography Development Programme in the Russian Federation for the period until 2020. This Programme provides framework for production of the retrospective bibliographic index "List of Books by the Indigenous Minorities of the Republic of Sakha (Yakutia): 1932–2014"; underway is the work on the retrospective bibliographic index "Union Catalogue of the Indigenous Minorities of the Russian North, Siberia and the Far East. 1812–2014".

Another important direction in the Centre's activity is cultural work and enlightenment among the population, as well as organizational and methodological work with the libraries of the republic and regions that participate in the Interregional Centre's projects. The most successful projects are the interregional library competition "The North in the DVD Format"

which was held in cooperation with the Russian Library Association, and the republican project "Between the Tundra and Taiga".

We have established contacts with the Japanese Ainu and Inner Mongolia Evenki for the purpose of replenishing information resources in the native languages of these peoples.

In October 2014, the National Library is holding the interregional scientific conference and workshop under the title "Book Culture of the North", and we kindly invite you to take part in it.

To conclude, we would like to note that the Interregional Information Centre of the Documentary Cultural Heritage of the Indigenous Minorities of the Russian North, Siberia and the Far East, which has been established under the National Library of the Republic of Sakha (Yakutia) contributes to the preservation of the documentary heritage of the indigenous peoples of the North and offers access to it through union information resources. This way the National Library of the Republic of Sakha (Yakutia) and the libraries of the northern regions of Russia support the development of linguistic and cultural diversity and representation of cultures and languages of the indigenous peoples of the North in cyberspace.

SECTION 4.
EDUCATION FOR PRESERVATION OF LINGUISTIC AND CULTURAL DIVERSITY IN CYBERSPACE

Farah MOTLAK

Deputy Minister of Education of Syria
(Damascus, Syria)

The Role of Education in Preserving Linguistic Diversity

Language is known as "a collection of signs with collective common indications that are possible to be spoken by speakers in community. It has a relative stability in every situation through which it appears / and it has a specific system through which it is consisted to construct more complex signs". These signs evoke the thing symbolized in the mind that knows its meaning, and vice versa, the thing provokes its symbol.

Language can be formulated through environmental and social factors in addition to ways of thinking prevailing in a society. At the same time, it can form the container or the outward appearance through which thinking is translated. Language is like a mirror which reflects thought because it is the most expressive accurate and comprehensive tool. You cannot speak about something that you cannot think about, because words are the linguistic expression of the corresponding ideas in the human's mind. On the other hand, you cannot think about things which you cannot express linguistically, and this confirms the presence of a dynamic reciprocal relationship between language and thinking. Each of them affects the other and is affected by it since we cannot talk about something that we cannot think about, and we cannot think of anything away from our linguistic capacity.

Tasks

Language reflects distinct aspects to perceive the world and it transfers among nations and generations, it is also a window for values expressing cultures and social life and it is a crucial factor in the formation of the identities of groups of people. However, more than 50% of the world's 6,000 languages face the danger of extinction, 96% of them are spoken by only 4% of the world's population, and less than a quarter of all languages is used in education. Besides, only about a hundred languages are regarded in education, in public life and in the digital world. The reinforcement of the diversity of languages and multilingualism,

particularly in education, culture, media and public life, is a prerequisite to ensure equal benefitting from education and knowledge and the probable equity of everyone's participation in human development and in general to ensure respect for the identity of each person and group.

Languages are tools for social development and communication, and we cannot do without them to transfer and develop knowledge; and due to their amazing diversity they constitute of evidences of humans' creative activity. Languages involve past experiences, cultures and identities, and at the same time carry aspirations and dreams of the future.

The Relationship between Linguistic and Cultural Diversity

Also, there is a great aware of the mutual integrated relation between the biological diversity and the cultural one. Due to the dangers that threaten the linguistic diversity, the Executive Board of UNESCO(at its 171st Session) for the year 2005 considered languages as a dimension located in the central point for all interactions with social and natural environment.

The relation between language and cultural identity is a strong inseparable one. In spite of the relation between language and culture, but the risk of prevailing of one particular language may lead to the obliteration of others languages and cultures, and this caution is not only found in the developing countries, but also in developed countries that have expressed more than once their awareness towards such a matter; and they worked to follow all the procedures to protect language and culture, especially in the field of television broadcasting and cultural and media industries.

The Importance of Linguistic and Cultural Diversity

UNESCO has recognized the seriousness of the demise of a lot of languages so it urges Member States and the Secretariat to promote the preservation and protection of all languages used by peoples of the world. UNESCO declared 2008 to be the International Year of Languages in pursuance of the 33rd session of the year 2005. Somehow, we find an organic link between language and culture. They are two containers of identity and they are at the same time the most important effecting factors in the area of influencing identity, communication, social integration, education and development. There has been a growing awareness of the vital role played by languages in development and in ensuring cultural diversity and intercultural dialogue, in achieving good quality of education for all people and strengthening cooperation in the construction of comprehensive knowledge societies, in preserving cultural heritage and stimulating policies of applying the benefits of science and technology in order to achieve sustainable

development. At this point, the urgent need to take action on the promotion of international commitment appears to promote multilingualism and linguistic diversity, which also include the protection of languages. Since language is a central issue in all the areas that gained the attention of UNESCO, the organization is promoting a multidisciplinary approach to multilingualism and linguistic diversity, which includes all programme sectors: education, culture, science, communication, information, and social sciences and humanities.

In addition to the importance of the mother tongue language, some linguists confirm the importance of cultural diversity in the areas of development, and not to be limited to the areas of values and morals. That is because of the importance of multilingualism in economic and productive activities and in the establishment of a democratic society and to promote education and respect for human rights. In addition to language there are those who emphasize the importance of culture in development where having culture at the heart of development policy is an essential investment in the future of the world and a prerequisite for successful globalization processes that take into account the principles of cultural diversity. Reminding most of the countries of this issue is the responsibility of UNESCO.

Some ongoing projects since the seventies have failed because they did not notice that evolution is not synonymous with economic growth alone, but it is a way of making intellectual, emotional, moral and spiritual life more perfect. So, it is not possible to separate between development and culture. Strengthening the contribution of culture and sustainable development was a goal which was launched within the framework of the World Decade for Cultural Development (1988–1998). Since then it has been significantly achieved due to a set of instruments for the standard-setting tools demonstration, such as cultural statistics, inventories and mapping of cultural resources on regional and national levels.

The Role of Education

The great challenge in this field lies in convincing political decision-makers and local and international social actors to integrate the principles of cultural diversity and values of multiculturalism in the overall policies, mechanisms and general practices especially through public and private companies. Preservation of language and cultural identity requires work to find a solution and here the role of education appears because it is the main way which communities follow in order to protect and confirm their cultural identity since the value of Man is what he achieves through his knowledge and the civilization of a society is the total sum of its members' knowledge which education donated them". Education plays a major role in supporting the values of loyalty and belonging and in building

Man's values and morals, and bringing him up within the light of pluralism which can develop the spirit of tolerance and the rejection of intolerance, the respect for others and the acceptance of differences with others. Proper bringing up obliges the individual to communicate and establish a mutual dialogue with others without hesitation. Through education cultural communication can be achieved through a set of procedures. Here are some of them:

- Confirming the importance of the concept and values of global citizenship, and benefiting from world education programmes to promote understanding and international cooperation among different peoples. It is one of the ways to face the challenges of the 21st century (working in mutual groups within the light of pluralism, and the equal dialogue among cultures, and Man's understanding of his role in the global system, and that he is one of its components and it requires his efforts and needs his positive participation, and realizing the overall picture of the global system, and the appreciation and respect for differences, regardless of gender and color), and this is indicated by UNESCO in its report on education in the 21st century.

- Emphasizing the importance of cultural diversity and multilingualism in the content of the curricula as well as teaching and learning activities.

- Providing learners with information and authentic experiences from other countries with their culture diversity.

- Showing interest in teaching foreign languages in addition to the mother tongue language and connecting schools to the Internet to make it easier for learners to have cultural communication and to create dialogue with others.

- Deepening cultural identity, strengthening and estimating people's participation through teacher's efforts.

In education, in the field of cultural and linguistic diversity, a teacher has to be careful to avoid three things when it comes to make judgments about some of the language or behavior patterns in a particular culture, the first thing: to contempt these behavioral patterns, since the behavior in any society has its justification and motive. Second: the application of the standards of other cultures since every culture has its standards and each language has its origins and foundations. Third: grabbing these types from their environment and looking at them abstracted from any link with the surrounding circumstances and from the characteristics of the society in which they were issued.

Susana FINQUELIEVICH

Director, Research Programme on Information Society,
Gino Germani Research Institute, University of Buenos Aires;
Principal Researcher,
National Council for Scientific and Technical Research
(Buenos Aires, Argentina)

Patricio FELDMAN

Member, Director, Research Programme on Information Society,
Gino Germani Research Institute, University of Buenos Aires;
Researcher, Civil Association for the Study
and Advancement of Information Society
(Buenos Aires, Argentina)

Celina FISCHNALLER

Research Assistant,
Research Programme on Information Society,
Gino Germani Research Institute, University of Buenos Aires, and
Civil Association for the Study and Advancement of Information Society
(Buenos Aires, Argentina)

Public Policies for Multilingual Education Using ICT in Latin America

Acknowledgments

Abstract

Many Latin American nations have committed themselves to becoming Knowledge Societies in the near future. They have approved development plans for horizons extended to 10, 15 or 25 years, with a view to substantially change their economies and their societies. The immediate implication is that most of their citizens will not just be connected to the Internet; they will have to be

qualified users and producers of ICT products and services, including contents, software, hardware, and new organizational patterns. The old "digital divide" related to devices and connectivity has been replaced with the new "knowledge divide", which is about people knowing how to use digital tools productively. In order to become real citizens of a knowledge society, the knowledge divide must be overcome. In Latin American countries, the language barrier between official languages, such as Spanish and Portuguese, and indigenous languages is an issue that still keeps many of these peoples from becoming productive cyber-citizens and taking advantage of universal access to information.

According to UNESCO, languages are powerful instruments for preserving and developing culture. Information and communication technologies (ICT) can help not only to encourage linguistic diversity and multilingual education but also to increase awareness and transmission of linguistic and cultural traditions throughout the world, and to motivate solidarity between diverse peoples. However, at present indigenous Latin American languages are not represented in the digital world. Language presence in cyberspace is insufficient in view of the increased importance of the role of cyberspace for access by indigenous peoples to education and information, the preservation and strengthening of their own languages and cultures, and the construction of inclusive knowledge societies.

This paper, based on meta research, focuses on Public Policies for Multilingual Education using ICT in Latin America. As we started our research, we had to decide the universe in which we would work. We chose to work on indigenous languages since this is widely considered as a vacancy area.

The paper starts with a panorama about multilingualism in Latin American countries. It analyzes national public policies regarding multilingualism, with the goal of introducing multilingualism to public education. The work also describes the interfaces between multilingual education and the new educational programmes using ICT.

Conclusions derived from the research are, among others, that Latin American multilingual education in cyberspace is acquiring an increasing importance due to the programmes of digital education and literacy, such as the plans of One Laptop Per Child. Increasingly these plans are including contents about indigenous languages and cultures. However, this tendency is recent. Impact on the educational community have not yet been studied in depth.

The inclusion of multilingual and multicultural contents in education, either at school or in cyberspace, has often been interpreted as proving a platform for oral traditional stories and local folkloric manifestations. However, bilingual intercultural education means much more than revaluation and dissemination

of folkloric displays. What is necessary for intercultural education in LA is to strengthen the cultural identity of indigenous peoples, not enhance confinement within their own traditions or facilitate the spread of their folklore, but to generate symmetric conditions of reciprocal interactions and exchange with the dominant culture.

Finally, the paper suggests measures to improve public policies and strategies regarding multilingual education in cyberspace for Latin American countries.

1. Introduction

The world is increasingly dependent on technological progress and the abilities to use these advances. Being able to access the Internet, either via computer or smartphone, is essential – not for personal communication and creativity of the individual, but also at the societal level for the delivery of services and as the foundation for education[152].

As we have stated in a previous work [Finquelievich and Bassi 2013], many developing nations have committed themselves to becoming Knowledge Societies in the near future. They have approved development plans for horizons extended to 10, 15 or 25 years, with a view to substantially change their economies and their societies. The immediate implication is that most of their inhabitants will have not only to be connected to the Internet and become qualified users of ICT, but also to be proactive citizens in the new Knowledge economy and society.

The old "digital divide" related to devices and connectivity has been replaced with the new "knowledge divide", which is about people knowing how to use digital tools productively. In order to become real citizens of a knowledge society, this knowledge divide must be overcome. The language barrier is an issue that still keeps many of these citizens from becoming productive cyber-citizens and enjoying universal access to information.

According to UNESCO, languages are powerful instruments for preserving and developing culture. New information and communication technologies (ICT) can help not only to encourage linguistic diversity and multilingual education but also to increase awareness and transmission of linguistic and cultural traditions throughout the world, and to motivate solidarity between diverse peoples. However, at present less than one hundred languages are represented in the digital world. Language presence in cyberspace is insufficient in view of the increased importance of the role of cyberspace for access to education

---

[152] *ICT in EDucation, UNESCO Bangkok, http://www.unescobkk.org/education/ict/online-resources/ databases/ict-in-education-database/item/article/digital-citizenship-in-a-cybersmart-world/.*

and information, and the construction of inclusive knowledge societies [Bassi and Finquelievich 2013]. Indigenous peoples need to access cyberspace and contribute their own world vision in their own languages so that they may participate in productive innovation processes, preserve indigenous languages and culture, and become full citizens.

This paper, based on meta research, focuses on Public Policies for Multilingual Education using ICT in Latin America. As we started our research, we decided to work on indigenous languages, since this may be considered as a vacancy area.

The findings we made confirmed this initial idea. Although there are an increasing number of studies about multilingualism in Latin America, they are mainly focused on the use of Spanish, Portuguese, French, and English [Prado & Pimienta 2012, among others]. Less attention has been paid to research on the role played by indigenous languages in cyberspace. While there is much literature about indigenous languages and ICT in Asia and Africa, much less is written about their counterparts in Latin America.

The scarce literature on this subject (even more scarce in English) reveals a vacancy area which needs development, as well as attention and technical support from UNESCO (through IFAP and the Permanent Forum on Indigenous issues (UNPFII), and other international organizations.

2. Multilingualism in Latin America

The "European conquest" of Latin America resulted in the dramatic reduction or disappearance of its indigenous peoples, languages and cultures. War, genocide, slavery, and disease reduced, absorbed or eliminated the native population. Native languages were mostly relinquished to the region's official languages of Spanish (Castilian) and Portuguese.

Argentina and Uruguay are extreme examples of this pattern of indigenous extinction.

According to the 2010 National Census in **Argentina**, only 955,032 people in a total population of 36,260,130 (2.6%) are or consider themselves to be indigenous or *mestizo* (mixed). While Spanish is the sole official language of the country, Mapudungun, Guaraní, and Quechua are still spoken in some provinces but less than Spanish, Italian, English, German and even French, Russian and Welsh. In the Provinces of Chaco and Corrientes, vernacular languages are recognized as official.

In **Uruguay,** the 2011 National Census revealed that only 4.9% of the population considered themselves descendants of the Charrúas, the original indigenous inhabitants. Again, Spanish is the only official language in Uruguay.

Other Latin American countries have populations that are largely indigenous or mestizo. Their indigenous languages are still widely spoken, and in many cases recognized as co-official along with Spanish.

In **Nicaragua**, while the official language of Spanish is broadly spoken (almost 95%, according to some sources), other *de facto* languages such as Creole, Miskitu, Rama language and Mayangna (Sumu) are widely used within their own linguistic communities.

In **Brazil**, Portuguese is the official language but more than 100 languages are spoken, mainly European and Asian in the urban areas and indigenous languages in the Amazon. According to estimates of the Socio-Environmental Institute (ISA) (Gaspar 2011) approximately 350,000 Indians currently live in communities throughout Brazil, with over 192,000 in urban centres. Although the official language of Brazil is Portuguese, there are about 180 indigenous languages spoken in the country – not counting the isolated Indians, who due to lack of contact with society have not yet been able to be met and studied (Gaspar 2011). Brazil is one of the 8 linguistically mega diverse countries in the world. The 1988 Constitution (Art. 210 and 231) recognizes the indigenous peoples' right to speak their own languages (Lei *de Diretrizes e Bases da Educação Nacional* 1996. Since 2003, together with Portuguese, the Ñe'engatu, the Tukano and the Baniwa have official language status. The 2010 census counted 305 different ethnics in Brazil, speaking 274 different languages.

Most Latin American countries have declared Spanish as their official language.

Chile, however, ratified their constitution in 2006 to permit official use of four indigenous languages in certain regions and communities: Aymara, Mapudungun, Quechua and Rapa Nui (Easter Island in Polynesia). In **Honduras** Afro-Caribbean English and indigenous languages are found in rural outskirts of the country. In **Colombia** Andean indigenous languages and Afro-Caribbean languages are spoken in the Choco region on the Pacific coast. In **Guatemala**, there are 23 distinct Mayan languages. Not all Guatemalans speak Spanish, and some only as a second or third language. In **Venezuela**, Afro-Caribbean dialects are found in the Caribbean and indigenous languages in the Guayana province.

While English is the official language of **Guyana**, parts of the population speak Hindi, Chinese, and a variety of indigenous languages. There is also a small Portuguese-speaking community.

A few countries are officially multilingual:

In **Paraguay,** 48% of its population is bilingual in both official languages, Guaraní and Spanish, with 37% speaking only Guaraní and 8% speaking only Spanish, but the latter increases with the use of Jopará, a colloquial form of Guaraní which uses large numbers of Spanish-related words.

Bolivia is officially multilingual, supporting Spanish and 36 native languages.

Ecuador defines Spanish as its official language, but Spanish, Quechua and Shuar are considered "official languages of intercultural relations" in Article 2 of their 2008 Constitution.

Peru has two official languages. The first official language is Spanish. Quechua, Aymara and other aboriginal languages also have official status in the zones where they are predominant. (Political Constitution, Art. 48). The most common languages are Spanish, to a lesser extent, Quechua and Aymara languages, not to mention numerous Amazonian languages, such as Urarina, Wampis, Shapra, Achuar, Shawi, and Awajún, among others

In **Mexico,** in addition to Spanish the government recognizes 62 indigenous languages, including Nahuatl which is spoken by more than 1.5 million people, and Aquacatec which is spoken by only 27 people. There is no official language at the federal level, although Spanish is the *de facto* state language.

In **Guatemala**, 23 indigenous languages are co-official with Spanish.

Vélez (2008) explains to us that Latin American multilingualism, which encompasses nearly half a million native languages as well as several foreign languages, is a product of African slavery plus European and Asian migration.

None the less, only a few countries in Latin America give official recognition to indigenous languages; most approach indigenous languages as minority languages of little importance. Intercultural Bilingual Education (IBE) is considered a tangent educational system, with specific civil servants or general officers in charge. The problem is that from its perspective there is no place for articulation between languages and cultures: each culture and language has its own field of action. Indigenous languages also occupy a disadvantaged position regarding school infrastructure and location, teacher training, etc. when compared to traditional education in Spanish and Portuguese.

However, indigenous languages, such as Aymara or Guarani, cover various countries, regardless of their national frontiers. Their number of speakers is far from being marginal. Aymara is spoken in Bolivia, Argentina, Chile, and Peru (total number of speakers in all countries is 2,589,000).

Guarani, specifically the primary variety known as **Paraguayan Guarani**, is an indigenous language of South America that belongs to the Tupí–Guaraní subfamily of the Tupian languages. It is one of the official languages of Paraguay (along with Spanish), where it is spoken by the majority of the population, and where half of the rural population is monolingual. It is spoken by communities in neighbouring countries, including parts of northeastern Argentina, southeastern Bolivia and southwestern Brazil, and is the second official language of the Argentine province of Corrientes since 2004; it is also an official language of Mercosur.

Guarani is one of the most-widely spoken indigenous languages of the Americas and the only one whose speakers include a large proportion of non-indigenous people. This is an anomaly in the Americas where language shift towards European colonial languages (in this case, the other official language of Spanish) has otherwise been a nearly universal cultural and identity marker of mestizos, and also of culturally assimilated, upwardly mobile Amerindian people[153].

Quechuan, also known as *runa simi* ("people's language"), is a Native South American language family spoken primarily in the Andes, derived from a common ancestral language[154]. It is the most widely spoken language family of the indigenous, with a total of probably some 8 million to 10 million speakers. Today, Quechua has the status of an official language in Bolivia, Ecuador and Peru, along with Spanish.

Currently, the main obstacle to the diffusion of the usage and teaching of Quechua is the lack of written material in the Quechua language, specifically books, newspapers, software, magazines, etc. Thus, Quechua, along with Aymara and the minor indigenous languages, remains essentially an oral language. In recent years, Quechua has been introduced in Intercultural bilingual education (IBE) in Bolivia, Ecuador and Peru, which is, however, reaching only a part of the Quechua-speaking population. There is an ongoing process of Quechua-speaking populations shifting to Spanish for the purposes of social advancement.[155]

3. Latin America Public Policies on Multilingualism in Education

Public policies on intercultural bilingual education were originated in the late 1960s, with the dissemination of a critical discourse, based on Paulo Freire's ideas, which questioned the official education that ignored the diversity of

---

[153] *http://en.wikipedia.org/wiki/Guarani_language.*

[154] *http://en.wikipedia.org/wiki/Quechuan_languages.*

[155] *http://en.wikipedia.org/wiki/Quechuan_languages.*

languages and cultures in Latin America. Since the 1970s the indigenous peoples started to claim the recognition of their cultural patrimony. They expressed the need to receive an education which included the contact between the multiple languages and cultures in the LA territory.

In 1983, in a UNESCO meeting about "The largest educational project in Latin America and the Caribbean" it was decided to replace the concept of biculturalism by interculturalism [Turbino 2004]. The issue was not teaching about two separate, unconnected cultures, but recognizing the process through which diverse cultures meet and nurture each other. Cultures were conceived as diachronic processes that develop and change with time and history, instead of synchronic entities which stay immutable through historic changes. Based on that conceptual change, Latin American countries started to implement bilingual intercultural education policies [Turbino 2004].

In **Argentina**, through the 2006 National Education Law, IBE is integrated to the Elementary, Primary and Secondary levels of education. It ensures the constitutional right of indigenous peoples to receive an education that preserves and strengthens their cultural inheritance, their languages, their cosmovision and ethnic identity; helps them to participate actively in a multicultural world, and improve their life quality[156]. The recognition of the IBE has initiated an institutionalization process over the entire national territory, in which several actors have actively participated: Education Ministries, Indigenous peoples, their organizations and communities, educational institutions, teachers and students. Although Argentina, as Uruguay, tended to promote assimilationist policies, in order to integrate every group into a national identity, including Spanish speaking. This national identity is usually more related to the European migration, than to the ancient indigenous peoples that have inhabited their territory for hundreds of years.

In recent years, **Bolivia**'s State has been restructured in terms of indigenous rights, defining itself as a plurinational state. The government is run by an Aymara President, Evo Morales. According to the Bolivian law, indigenous community is defined by self-recognition and good knowledge of native languages. The 2009 Constitution establishes Bolivia as a multilingual state which recognizes 36 different ethnic groups with their own languages or variations. Public service requires that all public employees speak at least two of the country's official languages. The Constitution was followed by a new Education Act that promotes the decolonization of education, aiming to strengthen the country's different cultures and languages (Danbolt 2011). However, bilingual education has been limited to the primary grades. The

---

[156] *http://www.me.gov.ar/consejo/resoluciones/res10/119-10_01.pdf*.

development of appropriate contents and pedagogical skills for teaching for IBE is still a challenge.

In **Brazil** an indigenous bilingual programme was institutionalized by the Presidential Act of 1966. The goal was to ensure bilingual education for indigenous groups and the right to maintain their languages. In 1973, the Brazilian government established a law to safeguard indigenous languages that required school instruction to be done in both the indigenous language and Portuguese. The Constitution of 1988 affirms the right of indigenous peoples to learn in their native languages and according to their own methods of learning. To that end, the Government started in 1991 a programme of "indigenous education" as a new model for intercultural and bilingual education with cross-cultural curricula aimed at strengthening culture, language, native teaching and learning processes and social infrastructure as a whole. There are currently 2.5 thousand indigenous schools in Brazil in 24 States of the Federation, attended by 177 thousand students. Between 2002 and 2007 the number of indigenous students grew at a rate of 45%. In the case of secondary education, there was a growth of over 600%. More than 90% of the 10 thousand teachers at indigenous schools of Brazil are themselves indigenous. The challenge now is to expand indigenous schools and the number of enrolled students. With regard to higher education, the Brazilian Government created affirmative action programmes to facilitate access by indigenous students to public and private universities across the country[157].

Since 1996, **Chile** has implemented the Bilingual Intercultural Education Programme (PEIB for its Spanish acronym). Its main goal is to "contribute to the development of the language and culture of original peoples, and to the training of intercultural citizens in the educational system[158]". PEIB's strategies include the teaching of indigenous languages, revitalization of original people's culture, intercultural education, and bilingualism. Since 2009 Chile has created studies programmes for indigenous languages such as Aymara, Quechua, Mapudungun and Rapa Nui, with the collaboration of indigenous peoples. In 2010, the National Council of Education approved the use of four native languages in the Basic Primary study programmes, which were implemented in 300 schools.

In **Colombia**, the new Constitution approved in 1991 recognizes a number of rights that specifically pertain to indigenous communities: The State recognizes and protects the ethnic and cultural diversity of the Colombian nation (Article 7);

---

[157]   *http://www.un.int/brazil/speech/11d-acs-ecosoc-Implementacao-Declaracao-das-Nacoes-Unidas-sobre-Direitos-Povos-Indigenas.html.*

[158] *http://www.mineduc.cl/index2.php?id_seccion=3442&id_portal=28&id_contenido=14010.*

it is the obligation of the State to protect cultural assets (Article 8). The languages and dialects of the ethnic groups are also official languages within their territories; in communities with their own linguistic tradition, education shall be bilingual (Article 10). Instruction shall respect and develop their cultural identity (Article 68)[159].

Moreover, the General Law of Colombian Education (chapter 3, Articles 55 to 63) recognized and institutionalized Ethno-education, which may be defined as "The education given to groups or communities which integrate the nation, and which have their own culture, language, tradition, and jurisdiction". Officially, ethno-education in Colombia has been understood as education from and for ethnic groups. Nevertheless, its emergence and implementation depend upon a request for greater autonomy by both indigenous communities and their leaders, so that they can make decisions on their own education, either under institutional boundaries or from the perspective of the communities' cultural projects [Mueses Delgado 2008]. In 2010, the Colombian State implemented the Law of Native Languages. Its goal is to grant recognition, protection, and development of individual and collective linguistic rights of ethnic groups which possess a linguistic tradition of their own[160]".

Besides state initiatives, the indigenous communities organized themselves in organizations such as the Indigenous Regional Council of the Cauca (CRIC), which has become an unavoidable example among the experiences of defense of cultural and linguistic legacy in Latin America. The creation of indigenous intercultural universities and research centres, as well as a collaborative self-diagnostic based on an ethno-educational model are a few examples of experiences that were implemented out of the State, but received support from governments, and allowed to make progress in the field of linguistic diversity (Murillo Mena 2011).

In **Ecuador** the rights of indigenous peoples and nationalities are contemplated in the National Constitution (Art. 57 concerning collective rights). The National State must guarantee the Bilingual Intercultural Education System (BIES), which will be used as the main means of instruction, the respective nationality's language, and Spanish as the language of intercultural relations. In March 2011, the Organic Law of Intercultural Education (OLIE) was approved. In their article 77, BIES has been recognized as a substantial part of national education through the Bilingual and Intercultural Education's Office. The BIES allows communities, peoples and nationalities to exercise

---

[159] *http://www1.umn.edu/humanrts/iachr/indig-col-ch11.html.*

[160] *http://www.lenguasdecolombia.gov.co/sites/lenguasdecolombia.gov.co/files/Ley_1381_2010_proteccion_ lenguas_nativas_0.pdf.*

collective rights, based on the intercultural, plurinational and multilingual state's character (Art. 78). Every teacher or school director has to speak and write in the respective nationality's language [Bacacela Gualan 2013].

In **Guatemala,** most of the population (50%) is from Mayan descent. Since the signing of the Peace Accords in December 1996, Guatemala has made significant advances in providing schooling for children at the primary level (grades 1–6). The Guatemalan Ministry of Education reports that the percentage of children completing their primary education has increased from 39% in the early 90's to 72.5% in 2006. However, a closer look at the data reveals a deep and ongoing disparity between the educational achievement and opportunities available for urban children of Latin descent as compared to children of Mayan descent living in rural areas. In addition, the disparity is amplified when comparing the education of boys and girls across all ethnic and socioeconomic factors[161]. Even if there is a State-implemented policy regarding the education of indigenous peoples, the most significant movements for linguistic diversity come from the indigenous people's initiatives, such as the Academy for Mayan Languages. The Academy regulates the use of the 22 Mayan languages spoken within the borders of the republic. It has expended particular efforts on standardizing the various writing systems used. Another of its functions is to promote Mayan culture, which it does by providing courses in the country's various Mayan languages and by training Spanish-Mayan interpreters[162].

México defines itself as a multicultural country. In Mexico, the main concern of intercultural education is tackling the relationship between the mestizo population and indigenous peoples. The descendants of ancient native people, heirs of their languages and cultures, have been viewed in various ways throughout history. One of these ways is to ignore their differences, grouping them as a whole in concepts like "Indian" or "Indigenous", referring to their ethnic origin, or putting them in terms such as "farmers", "laborers" or "migrants", related to the work they do or their need to move away from their homeland, and even concepts such as "Latinos", "Chicanos" or "Cholos" when they cross the border to the US, in reference to their Mexican origin [Bengochea 2010].

In 2003, the General Law of Linguistic Rights was promulgated. It declared both Spanish and indigenous languages as "national languages" in the whole Mexican territory. This law is related to the creation of the National Institute of Indigenous Languages (INALI). However, this legislation is not consolidated in practice. Many of the difficulties in the current multilingual education

---

[161] *http://www.avivara.org/aboutguatemala/educationinguatemala.html.*

[162] *http://en.wikipedia.org/wiki/Academia_de_Lenguas_Mayas_de_Guatemala.*

policies are linked to the lack of planning and training of the teachers charged to develop indigenous education, the insufficiency of learning methodologies and quality contents. Much of the education is still carried on in Spanish.

A relevant experience for Latin America in general is the Community University URACCAN[163] in **Nicaragua.** The University of the Autonomous Regions of the Nicaraguan Caribbean Coast (Spanish: *Universidad de las Regiones Autónomas de la Costa Caribe Nicaragüense*, abbreviated *URACCAN*), is a university founded in 1992. It is described as an "intercultural university community for indigenous peoples and ethnic communities"[164]**.** The university provides higher education to some of the country's most marginalized peoples, including the indigenous Miskitu, Mayanga, and Rama, and the Afro-Caribbean Creole and Garífuna, all of whom live in the eastern part of the country. While comprising only four percent of Nicaragua's total population, these coastal groups represent most of the country's cultural and linguistic diversity[165].

In **Perú,** Article 2 of the 1993 Political Constitution declares: *"The State recognizes and protects the ethnic and cultural plurality[166]"*. Besides, Article 17 claims that the State must encourage intercultural and bilingual education. Article 20 of the General Education Law sanctioned in 2003 states that intercultural and bilingual education must be provided in the whole education system. On July 5, 2011, the Peruvian Congress officially recognized indigenous languages by passing the Law for the Use, Preservation, Development, Revitalization, and Use of Indigenous Languages (Law 29735)[167]. Part of implementing international and domestic human rights legislation such as the UN Declaration on the Rights of Indigenous Peoples is respecting, protecting, and fulfilling the individual and collective right to speak one's native language. The law recognizes that language diversity is linked to the expression of individual and collective identity, as well as a different way of conceiving and describing reality, and that these languages should be celebrated as well as used nationally. It makes indigenous languages official languages of Peru. Public administration will now have to communicate in the 80 Indigenous languages spoken in Peru[168]. The law requires the Ministry of Education to conduct a

---

[163] *file:///C:/Users/Susana/Downloads/II_CAP_9.pdf.*

[164] *http://en.wikipedia.org/wiki/University_of_the_Autonomous_Regions_of_the_Nicaraguan_Caribbean_Coast.*

[165] *http://www.uvm.edu/sistercity/URACCAN.html.*

[166] *http://www.tc.gob.pe/constitucion.pdf.*

[167] *http://www.culturalsurvival.org/news/peru/peru-officially-recognizes-indigenous-languages.*

[168] *http://www.culturalsurvival.org/news/peru/peru-officially-recognizes-indigenous-languages#sthash.6i0X2OM7.dpuf.*

national register of indigenous languages and update Peru's ethnolinguistic map. This law repeals Decree Law 21,156, which recognized Quechua as an official language of Peru.

4. Multilingualism and Education in Cyberspace

"(…) Cyberspace presents both a threat and an opportunity for multilingualism", writes Prado [2012: 28]. He adds: "A threat, because the most highly equipped languages and those spoken in dominant states, impose themselves over others, and are supported by the network's technicality. An opportunity, because cyberspace's accessibility and universality allows it to give voice to languages that have been unable to make themselves heard via other recording and knowledge dissemination tools. We believe that this ease of access, Internet's ability to mobilize and coordinate many people, and its multimedia capabilities, will assure the rescue and revitalization of minority languages".

Since we're limited by time and space, we are focusing on Argentina, Mexico, and Peru, and specifically on their efforts for keeping alive multilingualism in cyberspace for educational purposes.

In **Argentina** the Law for National Education approved in 2006 created the Intercultural Bilingual Education (IBE) modality. According to this Law, the levels of Elementary, Primary and Secondary education must ensure the indigenous peoples inhabiting the Argentine territory the constitutional right to have access to an education which contributes to preserve their ethnic identity, their language, worldview, and culture.

In 2010 Argentina developed the National Programme "Conectar Igualdad". It aimed to achieve education, information and media literacy for the country's population. "Conectar Igualdad" grants democratic access to technological resources, reaching all public secondary schools in Argentina, both in urban and rural areas. The Programme is developed by the Argentine Republic Ministry of Education, the Social Security National Administration (ANSES), the Ministry for Federal Planning and Public Investment and Services, and the National Executive Cabinet s Head. Its original goal was to distribute 3 million netbooks to secondary school students and teachers, special schools, and institutes for teacher training. By May 2013 the Programme has delivered nearly 4 million notebooks. "Conectar Igualdad" focuses not just on the distribution of personal netbooks to teachers and students, but on other main goals as well: creation of a "technologic floor" that connects servers in order for all schools to have access to the Internet and creation of internal networks; generation of digital content; and development of a Federal training system for teachers on ICT-use in schools. "Conectar Igualdad" and the "Educ.ar"

platform, integrate the National Media and Information Literacy Campaign [Finquelievich, Feldman and Fischnaller 2012].

The National Programme of Intercultural Bilingual Education was created in 2006. Intercultural Bilingual Education was conceived as a strategy for educational equity, since it is based on the full participation of indigenous languages and cultures in the learning process. It recognizes sociocultural diversity as a positive asset for the society, promoting the development of rich and varied cultural traditions.

EIB's goals are: to design educational policies oriented to build an alternative approach of the sociocultural and sociolinguistic diversity in the Argentine educational system; promote jointly with indigenous peoples pedagogic strategies that respond to their specific needs and revert their historical exclusion from the educational system. EIB's Action lines are focused on teacher training, production of multilingual didactic content, and production of pedagogic projects, systematization of information about the educational situation of indigenous peoples, educational research, and grants for indigenous students. These goals are accomplished by special training given to indigenous and non-indigenous teachers and by the contents accessible by students and teachers at the national site Educ.ar.

In **Peru**, since 2008, the General Director of Education Technology at the National Ministry of Education has developed the programme "One Laptop per Child" aimed at delivering 600,000 computers to students and teachers of primary schools in rural, extremely poor communities. The programme's main objective is to reduce the huge gap between urban and rural schools, many of them located in remote areas, where one single teacher works with several school courses, lacking educational materials and access to technology. The Ministry of Education has distributed 513,204 computers and has trained over 5,144 teachers, and plans to extend the programme to secondary schools. The XO model computers provided to students can be taken home to be shared with families and friends, in order to socialize the computers' use and increase their impact on the communities.

The second stage of the programme seeks to improve the use of computers in urban areas where most people have a PC and can access connection. In this case, schools have several teachers and the use of XO laptops is intended to socialize. Equipment is delivered to each school and not to each student or teacher. For this purpose, Technology Resource Centres were created to share the use of machines and employ other technology resources such as mobile Internet, robotics, etc. Stage Three, implemented in 2011, seeks to extend the application of the high school programme, providing more than 600,000

laptops by the end of 2012. The characteristics of this phase are exactly the same as those of the second stage, but at secondary school level [Finquelievich, Feldman and Fischnaller 2012].

Computers and other ICT contribute to the improvement of quality, equity and relevance of intercultural bilingual education. Students will for instance use interactive bilingual modules to improve their learning in the classroom. This training will be part of the Peruvian Intercultural Bilingual Education (IBE). This means that students will be educated both in their mother tongue, Quechua, and in Spanish.

In **Mexico** the Secretariat for Communications and Transportation (SCT), the Ministry of Education, and the Ministry for Social Development are implementing three programmes intended to reduce the digital gap and take it to OECD levels in year 2015:

a) Habilidades Digitales para Todos (Digital Skills for All, 2010–2012) addresses primary school student use of ICT in the learning process, and the development of digital skills.

b) Campaña Nacional de Inclusión Digital Vasconcelos 2.0, (Vaconcelos National Campaign for e-inclusion) is to mobilize young students who are already skilled in ICT use to reduce the digital gap within the socially vulnerable adults group.

c) Centros Digitales Comunitarios E-méxico (E-Mexico Community Digital Centres, CCD) implements digital community centres in rural areas. E-México has created more than 3200 CCDs throughout the country, where people can have free access to the Internet. [Finquelievich, Feldman and Fischnaller 2012]

Moreover, the Digital Acquis of Indigenous Languages (ADLI)[169] produces didactic materials and contents which allow gathering, storing and systematizing information about indigenous languages in Mexico. Its main goal is to revert the threats undergone by these languages to revitalize the great Mexican intangible inheritance. Supported by the Max Planck Foundation, ADLI has a digital acquis capable of manipulating large quantities of text, audio and video for diverse purposes, such as linguistic analysis, edition and access to educational contents, and production of useful contents for the defense of the Mexican immaterial heritage. ADLI manages Internet portals, multimedia contents, and the production of books for children.

---

[169] *http://lenguasindigenas.mx/acerca-del-acervo-digital.html.*

All these countries, together with initiatives from Ecuador, Chile, Colombia, Venezuela and others, are integrating educational efforts with digital technologies in order to reach the overall population, including small and remote communities. It is interesting to remark that policies to integrate multilingualism and ICTs in education do not seem to be related with the proportion of indigenous peoples in diverse countries. For example, Argentina and Bolivia, which seem to be in extreme situations regarding the percentage of indigenous populations in their respective territories, are making similar efforts to implement multilingual ways of study.

Nevertheless, it appears evident that efforts for multilingual education in cyberspace cannot come only from State policies. It would be desirable to enhance bottom-up initiatives from indigenous communities, facilitating a constructive dialogue between governments and civil society, particularly the representative organizations of diverse ethnic and cultural groups.

Latin American Network of Education Portals (RELPE – Red Latinoamericana Portales Educativos)[170] offers an open access search engine for contents within Education Portals of Latin America and the Caribbean. The initiative to develop a collaborative network of education portals in Latin America began in 2001 within the framework of bilateral cooperation agreements made by several countries of the region. Education Portals are full members designated as such by the respective Ministry of Education in each country. It is required to have completed the protocol indexing web content and technical adjustments made to connect virtually to the Network.

In 2006, the International Development Research Centre of Canada (IDRC) approved the Draft Strategy for the Consolidation and Integration of the Latin American Network of Educational Portals (RELPE). The Executive Secretariat of RELPE is currently the responsibility of Argentina. RELPE gathers and makes accessible multicultural and multilingual contents, resources, and dictionaries to be used both in in-person classrooms and in virtual education[171].

5. Conclusions and Proposals

Conclusions

The existence of public policies for intercultural and multilingual education in cyberspace in LA countries does not necessarily have a direct relation with

---

[170] *http://www.unesco.org/new/en/communication-and-information/portals-and-platforms/goap/key-organizations/latin-america-and-the-caribbean/relpe/.*

[171] *http://relpe.org/multicultural/.*

the number of indigenous peoples and languages in the countries. While some countries such as Argentina, where the indigenous population counts less than 3%, are developing effective public policies with the participation of indigenous peoples, in other countries with a considerable proportion of indigenous population policies and actions do not come primarily from the National State, but from indigenous organizations.

Indigenous social movements have obtained some success, achieving the recognition of indigenous rights by the National states. However, in many cases, the new national and international legislation regarding indigenous rights and languages does not go further than well intentioned declarations. One of the most evident problems in LA is not the lack of good legislations regarding multilingual rights, but the lack or insufficiency of policies' implementation, as a consequence of deep-rooted discriminatory practices. Overcoming these limitations would require more participative and democratic consultation policies with indigenous peoples.

Multilingual education in cyberspace is acquiring an increasing importance due to the programmes of digital education and literacy, as the One Laptop Per Child plans. Increasingly these plans are including contents about indigenous languages and cultures. However, this tendency is recent. The impacts on the educational community have not been studied in depth yet.

The inclusion of multilingual and multicultural contents in education, either at school or in cyberspace, has been interpreted often as giving place to oral traditional stories and to local folkloric manifestations. However, bilingual intercultural education means much more than the revaluation and dissemination of folkloric displays. What is necessary for intercultural education in LA is to strengthen the cultural identity of indigenous peoples, not to enhance their confinement in their own traditions, or to facilitate to better sell their folklore, but to generate symmetric conditions of reciprocal interactions and exchange with the dominant culture.

Cultural diversity is not limited to rural areas. When planning multicultural and multilingual education in cyberspace it is necessary to consider BIE not only for rural areas, but also for urban and urban marginalized areas. BIE in urban marginalized areas is the new challenge in LA cities.

Intercultural and multilingual diversity should not be limited to primary education. In LA countries young indigenous students are accessing higher education. New indigenous and intercultural universities are being created. BIE should not only be "for all", but also "for lifelong education and training".

Proposals

1. *Using the educational system to support multilingualism and culture preservation*

Education plays a substantial role in any middle- or long-term effort to produce significant social and economic changes. And the educational system is in general represented in all communities, even the smallest ones. As such, it can become the spearhead for government actions destined to implement access and multilingualism policies. Therefore government actions through education can play a substantial role in enhancing knowledge access for its citizens by including the following strategies as part of their action plans:

- Fighting the access divide by providing free or low cost Internet access in schools, libraries and community centres. In areas where Internet access is still very expensive or cost of equipment prohibitive, school computer equipment could be used to offer community access in the evenings, and use them for training, access to government online portals (e-government), telework, community projects and supporting women and girls, etc.

- Supporting the digitization and preservation of content with anthropological or historical value: According to UNESCO: "*Cultural heritage is not limited to material manifestations, such as monuments and objects that have been preserved over time. This notion also encompasses living expressions and the traditions that countless groups and communities worldwide have inherited from their ancestors and transmit to their descendants, in most cases orally.*"

- Small communities are in many cases holders of valuable cultural treasures that will be lost forever unless they are documented. Schools can play a crucial role in detecting and digitizing songs, stories, customs, techniques, art, music, local languages, flora and fauna and other content of immeasurable cultural value. Governments can help by providing schools with digitizing equipment (cameras, scanners, recorders) and training teachers and librarians on digital preservation software. Through competitions schools could collect, classify, store and share with the world valuable content that would otherwise be lost forever.

- All kinds of existing technologies should be considered for multilingual use and education in cyberspace: mobile cell phones, and portable digital audio (MP3) players, palm computers, pagers and handheld digital games. This communication revolution will impact global language diversity. Already in the near future, we may witness divergent

trends of multilingual usage in cyberspace with novel possibilities for languages, big and small alike [Ikvovic and Lotherington 2009].

- Building a multicultural pedagogic proposal in all educational stages, at national and regional levels. More actions are needed, such as the implementation of an Observatory of Latin American linguistic and cultural diversity, with the direct participation of indigenous communities. Recent initiatives, such as creating indigenous universities (i.e. the Intercultural Indigenous University (UII) in México, the Autonomous Intercultural Indigenous University[172], or the Universidad Autónoma Indígena Intercultural (UAII) in Colombia are initiatives which should be strongly supported to promote new articulations among indigenous peoples and the academia.

- Giving online counseling to indigenous teachers and professors in conflictive situations with the national and regional educational systems.

- Supporting the organizations of indigenous educators in cultural and educational activities (training of indigenous educators, congresses, virtual forums, etc.).

- Using the Internet to gather updated information about schools of indigenous modality, number of indigenous and non-indigenous educators and students, etc.

- Strengthening regional educational portals such as RELPE to disseminate multilingual and multicultural educational contents.

2. Promoting cultural diversity in the digital world by:

- Encouraging the creation and processing of and access to educational, cultural and scientific content in digital form in schools, in order to ensure that all cultures can express themselves and have access to the Internet in all languages, including indigenous ones.

- Fostering the creation and dissemination of content in local languages and promoting through the educational system mechanisms for the production and distribution of user and community generated content, thereby facilitating communities' access to contents in their own languages.

- Supporting capacity building for the production of local and indigenous content on the Internet. School based Telecentres can be useful for

---

[172] *http://www.cric-colombia.org/portal/universidad-autonoma-indigena-intercultural-uaii/.*

this purpose; school teachers, computer technicians and librarians can be trained to become digitizing experts.

- Promoting multilingualism on the Internet so that everyone can have access to the most critical content in their own language. For example, by developing electronic translators, dictionaries and language tools for indigenous languages, supporting translation of useful software tools, offering content in government sites in several languages, offering tax incentives and subsidies for the development of content and software tools in local languages, among other possible actions. Schools can play a crucial role in acting as mediators in all these activities. [Finquelievich and Bassi 2013]

- Promoting and supporting plurinational education portals, such as RELPE – Red Latinoamericana Portales Educativos[173], in order to provide access to search engines for multicultural and multilingual contents within Education Portals of Latin America and the Caribbean.

References

1. Bacacela Gualán, S. P. (2013). Informe del sistema de educación intercultural bilingüe en Ecuador, https://docs.google.com/document/d/1v-sjXL1H5D2_dxMjopV21umLxuIztHpsuEw0lrwqMdc/edit#.

2. Coppi Agostinelli, M. A. (2012). "Education inequalities of indigenous peoples in Ecuador and in Peru". Paper presented at the annual meeting of the 56th Annual Conference of the Comparative and International Education Society, Caribe Hilton, San Juan, Puerto Rico, Apr 22, 2012. http://citation.allacademic.com/meta/p552127_index.html.

3. Cummings, S. M. and Tamayo, S. (1994). Language and Education in Latin America: An Overview. http://faculty.smu.edu/rkemper/anth_6306/anth_6306_language_and_education_in_latin_america.htm.

4. Danbolt, L. D. (2011). The challenge of bilingualism in a multilingual society: The Bolivian Case. NLA, School of Religion, Education and Intercultural Studies, Bergen, Norway, Journal of Intercultural Communication, ISSN 1404-1634, issue 27, November 2011. http://www.immi.se/intercultural/nr27/drange-27.htm.

---

[173]   *http://www.unesco.org/new/en/communication-and-information/portals-and-platforms/goap/key-organizations/latin-america-and-the-caribbean/relpe/.*

1. De Bengochea, N. Intercultural education and indigenous education in Mexico, an experience in Oaxaca. http://tsg.icme11.org/document/get/805.

2. Finquelievich, S. and Bassi, R. (2013). "The Role of Language in Knowledge Society Educational Systems", Educational Technology Debate. Exploring ICT and Learning in Developing Countries. https://edutechdebate.org/cultural-heritage-and-role-of-education/the-role-of-launguage-in-knowledge-society-educational-systems/.

3. Finquelievich, S., Feldman, P., Fischnaller, C. (2012). "Public Policies on media and information literacy and education in Latin America: Overview and proposals". Paper presented at the International Conference «Media and Information Literacy in Knowledge Societies», IFAP UNESCO, 24–28 June 2012, Moscow.

4. Flores Farfán, J. A. (2010). En torno a la política y planeación lingüísticas en el contexto Latinoamericano. Linguapax – Latin America Delegation Professor of Linguistics, CiESAS, Mexico D. F. http://jaf.lenguasindigenas.mx/docs/2010-entorno-a-la-planificacion-linguistica-en-anual-review-linguapax.pdf.

5. Gaspar, L. (2011): Indigenous Language in Brazil. Fundacao Joaquim Nabuco, Recife, http://basilio.fundaj.gov.br/pesquisaescolar_en/index.php?option=com_content&id=1122:indigenous-language-in-brazil.

6. Godenzzi, J. C. Globalización, Multilingüismo y Educación. El caso del Perú, Aportes Andinos N. 11, Aportes sobre diversidad, diferencia e identidad, Universidad Andina Simón Bolívar, Peru. http://red.pucp.edu.pe/ridei/libros/globalizacion-multilinguismo-y-educacion-el-caso-del-peru/.

7. Internacional de la Educación para América Latina (2014). Intercultural Multilingual Education in Latin America. Norway. http://www.ei-ie-al.org/publicaciones/indigenous_web.pdf.

8. Ivkovic, D. and Lotherington, H. (2009). Multilingualism in cyberspace: conceptualising the virtual linguistic landscape, International Journal of Multilingualism, Vol. 6, No. 1, February 2009. Routledge. file:///C:/Users/Usuario/Downloads/conceptualizing_VLL-3%20copy.pdf.

9. Lippenholtz, B., and Marés, L. (2013). El uso de las TIC en la preservación de las lenguas originarias de Latinoamérica, Educational Technology Debate, https://edutechdebate.org/cultural-heritage-and-

role-of-education/el-uso-de-las-tic-en-la-preservacion-de-las-lenguas-originarias-de-latinoamerica/.

10. Luna, J. (2012). Los charrúas: genocidio y resistencia, Voltairenet.org, http://www.voltairenet.org/article176599.html.

11. Monsonyi E., Rengifo, F. (1983). "Fundamentos teóricos y programáticos de la educación intercultural bilingüe". In: E. Rodriguez Masferrer y R. Vargas (eds.), *Educación, etnias y descolonizaciónn en América Latina*. México, UNESCO III, Vol. 1, pp. 209-230.

12. Mueses Delgado, C. A. (2008). Etnoeducación en Amazonas: ¿se indianizó la institucionalidad educativa o se institucionalizó la propuesta indígena?, Revista Educación y Pedagogía, vol. XX, núm. 52, Septiembre–Dec. 2008, http://aprendeenlinea.udea.edu.co/revistas/index.php/revistaeyp/article/viewFile/9881/9078.

13. Murillo Mena, M. E. Sentidos de formación de maestros y etnoeducadores en la Universidad Tecnológica del Chocó – UTC H – Colombia. ARJÉ Revista de Postgrado FACE-UC. Vol. 5 Nº 8. Enero-Junio 2011, 27-47. http://servicio.bc.uc.edu.ve/educacion/arje/arj08/art02.pdf.

14. Ocampo, G. (2013). La CTERA en la Argentina. In: III Reunión Regional Educación Pública y Pueblos Indígenas, 15, 16 March 2013. Cuzco, Peru.

15. Orán, R., Orán, K. Y., Ologwagdi, Wagua, A. (Equipo EBI Kuna) (2009). Educación Bilingüe Intercultural en los Territorios Kunas de Panamá, Congreso General Kuna, Congreso General Kuna de la Cultura.

16. Prado, D. (2012): A brief return to the genesis of Net.lang. In: L. Vannini and H. Le Crosnier (eds.). Net.lang. Towards the multilingual cyberspace. C&F Editions, Cen, France, http://net-lang.net//externDisplayer/displayExtern/_path_/netlang_EN_pdfedition.pdf.

17. Prado, D. & Pimienta, D. (2012). Public Policies for Languages in Cyberspace. In: L. Vannini and H. Le Crosnier (eds.). Net.lang. Towards the multilingual cyberspace. C&F Editions, Cen, France, http://net-lang.net//externDisplayer/displayExtern/_path_/netlang_EN_pdfedition.pdf.

18. Tubino, F. (2005). La praxis de la interculturalidad en los Estados Nacionales Latinoamericanos. Cuadernos Interculturales, vol. 3, núm. 5, julio-diciembre 2005. Universidad de Valparaíso, Chile.

19. Tubino, F. El interculturalismo latinoamericano y los Estados Nacionales. Available at: http://www.cajanegra.buap.mx/109.pdf.

20. UNESCO (2012). ICT in education in Latin America and the Caribbean. A regional analysis of ICT integration and e-readiness. Montreal, Quebec, Canada, http://unesdoc.unesco.org/images/0021/002179/217983e.pdf.

21. UNICEF (2009). Atlas Socio-lingüistico de Pueblos Indígenas en América Latina. Cochabamba, Bolivia, http://www.unicef.org/honduras/tomo_1_atlas.pdf.

22. Vannini, L. and Le Crosnier, H. (2012). Net.lang. Towards the multilingual cyberspace. C&F Editions, Cen, France, http://net-lang.net//externDisplayer/displayExtern/_path_/netlang_EN_pdfedition.pdf.

23. Wagua, A. (2014). EBI Guna. Nan Gaburba Oduloged Igar. Génesis, enfoque, estrategias y métodos aplicados. Gunyala: Proyecto EBI-Guna. http://www.gunayala.org.pa/.

need of a new market for publicity, video on demand, music and other personalized services has appeared. In parallel, the development of tools and applications for networked systems enabling information diffusion at low cost to a great number of users has brought new opportunities to develop related services. In this context, the convergence between telecommunication, broadcasting and information technologies offered new possibilities for the design of services providing access to multimedia resources. This has generated a new branch of economy.

The number of people involved in this new sector tends to increase. Those professionals provide services such as localization of products and services, their promotion via the web, and online commerce. In this context, the knowledge of foreign languages is essential.

The degree in Applied Foreign Languages for Multilingual Management of Information Society (LEA-MSI) appears within this context. It aims to form professionals able to work in this new sector of the economy.

The LEA-MSI is a degree that was conceived within the scope of the French programme of Applied Foreign Languages (Langues Etrangères Appliquées, LEA). This programme exists in many French universities. The University of Brasilia has established partnerships with some of these universities in order to create a similar programme in Brazil.

Objectives

The main objective of the LEA-MSI is to form professionals with a solid knowledge of two foreign languages and the instruments that allow the management of multilingual information in the information society.

The degree aims to provide the students with competences in the following areas:

- foreign languages;
- information and communication technologies;
- management and economy;
- law.

384

need of a new market for publicity, video on demand, music and other personalized services has appeared. In parallel, the development of tools and applications for networked systems enabling information diffusion at low cost to a great number of users has brought new opportunities to develop related services. In this context, the convergence between telecommunication, broadcasting and information technologies offered new possibilities for the design of services providing access to multimedia resources. This has generated a new branch of economy.

The number of people involved in this new sector tends to increase. Those professionals provide services such as localization of products and services, their promotion via the web, and online commerce. In this context, the knowledge of foreign languages is essential.

The degree in Applied Foreign Languages for Multilingual Management of Information Society (LEA-MSI) appears within this context. It aims to form professionals able to work in this new sector of the economy.

The LEA-MSI is a degree that was conceived within the scope of the French programme of Applied Foreign Languages (Langues Etrangères Appliquées, LEA). This programme exists in many French universities. The University of Brasilia has established partnerships with some of these universities in order to create a similar programme in Brazil.

Objectives

The main objective of the LEA-MSI is to form professionals with a solid knowledge of two foreign languages and the instruments that allow the management of multilingual information in the information society.

The degree aims to provide the students with competences in the following areas:

- foreign languages;
- information and communication technologies;
- management and economy;
- law.

evolution of languages in cyberspace as well as the launching of a great number of multilingual international events has encouraged university centres and international education institutions to develop human resources for these new working market demands.

In a very dynamic working professional environment, a high number of activities nowadays require expertise both in foreign languages and a field of knowledge in which these languages can be applied. These activities are concentrated in institutions and international organizations such as: national and international public offices, foreign cooperation agencies, United Nations agencies (such as UNESCO, for example), university centres, state institutions such as Deputy Chamber and Senate, NGOs, multilingual libraries, particularly digital libraries and national and international press houses. A LEA-MSI professional will be equally able of assuming functions related to logistic support for international events, such as the important international sport competitions (on-going FIFA World Cup, Olympic Games, Universíade) taking place in Brazil in 2014, 2016 and 2017.

Given this context and taking into consideration the development of such activities related to multiculturalism in cyberspace, as a consequence of globalisation, there is a need of new professionals able to work within the framework of a new technological environment. The local context in the University of Brasilia, where over the years there are B. Sc programmes on international relations, economy, tourism and management has conducted the B. Sc in LEA to fill in the above gaps. As we will discuss later, the Bachelor in Applied Foreign Languages is prepared to work in interdisciplinary synergy with other areas, being able to act in both public and private sector, national and international and the services sector.

The Three LEA-MSI Application Areas: Audio-Visual, Terminology and Multilingualism in Cyberspace

The new B. Sc was launched in 2010. UnB became the 1st Brazilian educational institution offering a multidisciplinary study programme in the field of information society. Students will receive a solid background in foreign languages (they select two out of English, Spanish and French) reaching a total of 1,800 hours in classroom. In addition, an amount of 255 hours of disciplines belonging to the application axes: audio-visual, terminology and multilingualism in cyberspace. It is also mandatory to pursue a stage in a public or private institution or in NGOs. For further details, please visit LEA webpage at http://www.let.unb.br.

3. Applied Terminology in Cyberspace

The Relevance of Terminology

I will comment a bit on the logical framework of our B. Sc in Applied Foreign Languages. In addition to the linguistic expertise earlier mentioned, the applied component is constituted of "three thematic complimentary axes":

a) Multilingualism in cyberspace focuses on the infrastructure requirements for inclusion and linguistic vitality in cyberspace. We can simplify this concept by saying "from oral language to languages websites": methodologies, scripts, technological and information technology tools, measurement systems and linguistic policies belong to this thematic axis.

b) The audio-visual thematic axis covers sub-titling techniques, audio-description, audio-visual translation and all technical aspects related to access to information in a foreign language, including by people with special needs;

c) The terminology thematic axis corresponds to lexical studies, corpus linguistics, multilingual dictionaries, etc.

We can affirm, therefore, that in the "technological iceberg" for dealing with languages in cyberspace the audio-visual component corresponds to the visible component, the summit of the iceberg, the technological infrastructure for multilingualism is the invisible bottom of the iceberg. Terminology helps in making the liaison between the infrastructure and audio-visual.

In the context of a multilingual knowledge society, terminology is a multidimensional subject. It depends both on the approach and technical affiliation of players.

A recent study in the framework of the European project POINTER considers terminology as a many-faceted subject being, depending on the perspective from which it is approached and the affiliations of the person discussing it:

- a resource,
- a set of methodologies and procedures to be used in creating this resource,
- a factor in communication,
- a community of actors, and
- an academic discipline.

Terminological resources are also valuable in many other ways: as collections of names or other representations, as the object of standardisation and harmonisation activities, and as the input (or output) of a wide range of applications and disciplines, whether human or machine-based (see the Figure below). The range of applications to which terminology is of direct relevance was a primary motivating factor at the inception of the POINTER Project with its brief to analyse the situation of terminology in Europe, and to make concrete suggestions for a future infrastructure and activities.
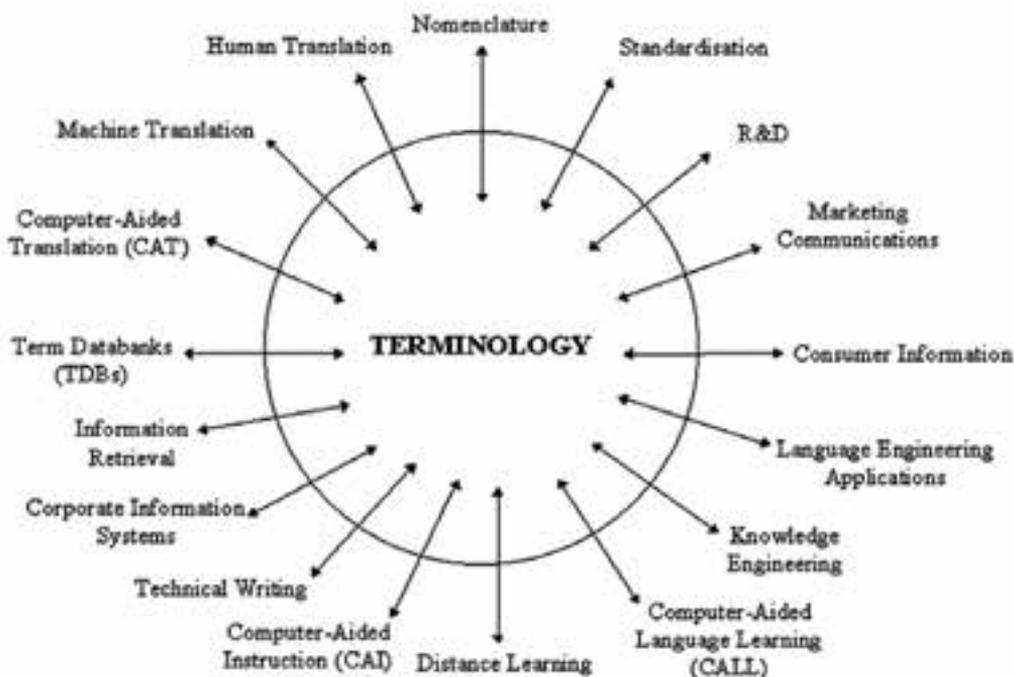

Figure 1. *Terminology Applications and Products*

Accelerated technological development and the emergence of new research areas and industries brings our society to shorter innovation cycles, an exponential growth of knowledge and the need to communicate even faster. There are studies indicating that the total amount of knowledge is doubling each 5 to 15 years depending on the area concerned.
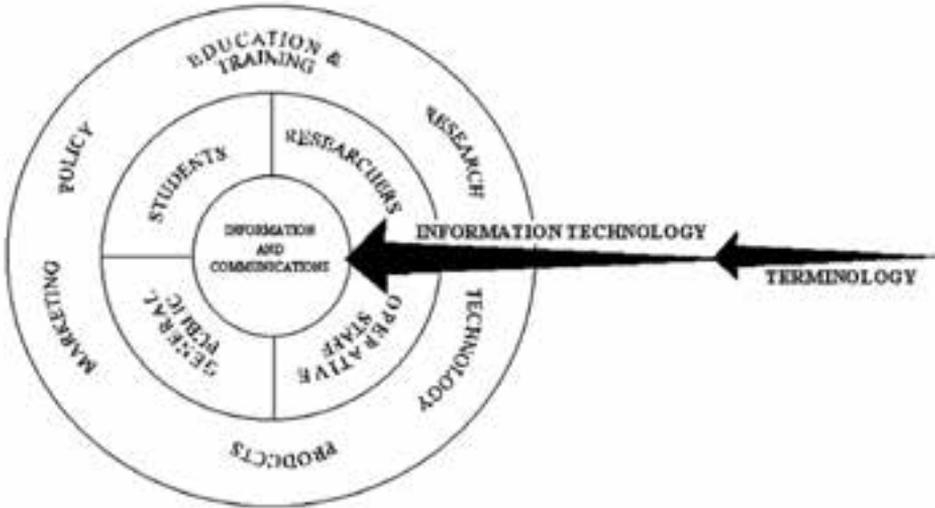
Figure 2. *Terminology: a key discipline for the Information Society*

The number of technical terms in each language highly developed is estimated at 50 million.

The POINTER Project also makes the following remark as regards multilingual terminology:

A point to be remembered here is that specialist (and indeed general) communication is normally an iterative and multilinear process, since knowledge is generally created in an evolutionary process and in several different places at once. Thus potential sources of uncertainty and misunderstanding arise in the form of homonyms (i.e. words that are used to denote more than one concept) and synonyms (i.e. more than one word for the same concept). This problem is becoming particularly acute with the strong tendency to interdisciplinarity in important modern scientific disciplines such as biotechnology, environmental science and materials science (it is a paradox that in this age of increasing specialisation science is becoming more and more interdisciplinary). At the same time, the risks involved in failing to communicate unambiguously and in a timely manner have often increased dramatically (two classic examples of this are the aerospace and environmental industries).

For all these reasons, contents-based information management is a prerequisite for improving the efficiency of communication. In addition, it should be borne in mind that communication is not solely monolingual, especially not within

Europe. In fact, there is a clear trend at the moment towards an increased awareness of multilingual issues, despite the predominance or at least lead function of English in the technical, business, economic, political and – to a lesser extent – cultural fields.

The Situation in Brazil

The importance of terminology in a multilingual information society has commenced to be present in Brazil, this being determined by the need of elaboration of technical manuals and usage in several languages. Meanwhile, this terminological multilingualism is still a little incipient and reaches only a few domains related to goods and services exportation. Some technical works about terminology in the information society have been recently published. This is the case of "Dicionário Brasileiro de Terminologia Arquivistica", a multilingual document which contains each term in German, Spanish, French, English, Italian and Portuguese. A regular work undertaken by the Brazilian Association of Technical Standards (Associação Brasileira de Normas Técnicas – ABNT) yields technical documents on terminology for industry.

4. Multilingualism in Cyberspace

Until 2003, the international concerns on languages were focused in ensuring the survival of endangered languages. At the 32nd UNESCO General Conference, the "*Recommendation concerning the Promotion and Use of Multilingualism and Universal Access to Cyberspace*" was approved. Such a Recommendation brings to the scene some principles and proposals towards a larger access, vitality and languages inclusion in cyberspace. At Brazilian national level, in addition to academic groups[174], in 2010 the "National Inventory of Linguistic Diversity" (INDL)[175] was created, estimating the existence of approximately 210 languages in Brazil, including indigenous groups as immigrants.

In this context, UnB's B. Sc offers a primer approach focused on three essential topics:

a) technology for inclusion of languages in the digital world;

b) measurement techniques of situation of languages in cyberspace;

c) linguistic policies.

---

[174] *In UnB, see http://www.laliunb.com.br/.*

[175] *http://portal.iphan.gov.br/portal/montarDetalheConteudo.do?id=15772&sigla=Noticia&retorno=detalheNoticia.*

5. Conclusions, Evaluation, Perspectives

This paper aims to:

a) explain the adopted rationale by UnB to carry out its education programme in Applied Foreign Languages to Multilingualism and Information Society;

b) present some aspects of the thematic axis, particularly terminology and multilingualism in cyberspace, in a B. Sc curriculum oriented towards a multilingual information and knowledge society.

While it is a bit premature to have an overall evaluation of the educational experience with this new B. Sc, some indicators of demand for the new programme (ratio candidates/offered places, jobs offer) allow to consider it as a consolidated study programme in UnB. Some academic final projects prepared by our students are also very much encouraging.

As regards the pertinence and the importance of terminology in information and knowledge societies, particularly in the industrial branches, the presented domains show a number of applications and products that constitute a permanent source of terminology use, in its unquestionable importance for multilingual information and knowledge societies.

Multilingualism in cyberspace is still a large avenue of knowledge for development, as we can confirm in the discussions here at this 3rd International Conference on Linguistic and Cultural Diversity in Cyberspace. Even an optimistic estimation based on the UNICODE Consortium tables shows 657 languages associated to digitalized scripts available. Methodologies, technologies and linguistic policies, as stressed in many international documents such as the Lena Declaration, will have to be further developed to ensure a truly multilingual cyberspace. Information for all will not become a reality, if educational and international institutions do not face the challenge of a multilingual cyberspace.

Irene KÄOSAAR

Head, General Education Department,
Ministry of Education and Research
(Tallinn, Estonia)

Minority Languages and Digital Environment: Friends or Enemies?

Preservation of the languages and cultures of minority peoples is a most important issue in the modern globalized world. The so-called international communication languages, English in the first place, are capturing ever more power. Proceeding from the current changes in the world, where communication is gradually shifting to the Internet and where the boarders of towns, countries and continents are no longer as significant as they used to be several dozens of years ago, the number of native speakers of a language and their ability to develop this language in the digital format are becoming crucial.

The population of Estonia is 1.3 million people. 70% of them are Estonians, about 25% are Russians, and the remaining 5% are people speaking approx. 120 different languages. The official language of Estonia is Estonian but historically Russian has also been used as a language of communication. Among the above 5% the largest national groups are Ukrainians, Finns, Belarusians and Armenians. Increasing is the number of residents whose primary language of communication is English.

Compulsory education in Estonia envisages 9 grades within which the medium of teaching may be chosen by the school owner, i.e., as a rule, by a local self-administration. As a result, in reality, we have mainly Estonian- and Russian-medium schools, the latter teaching some 20% of the total number of students. There is also a possibility of using English as a medium within the International Baccalaureate (IB) system, German at the senior high school level and Finnish at a small Finnish school.

The Estonian government pays great attention to weekend schools of the country's national minorities and supports them both financially and by human resources to stimulate their activity aimed at supporting the national minorities in learning and preserving their languages and cultures. This academic year, for example, the government grants have been allocated to 18 weekend schools of the national minorities. By doing this, the Estonian government ensures that the Azerbaijanian, Armenian, Uzbek, Tatar, Russian, Ukrainian, Belarusian, Korean and Ingrian languages are preserved in Estonia.

Based on the above, the government of Estonia is facing two tasks:

1) To maintain the development and attractiveness of the Estonian language in the digital environment, and

2) To turn the Estonian-language Internet environment into a common communication space for all Estonian residents. We consider it important to provide for all residents a common information environment, an environment which is permanently attractive, educational and affordable.

Estonia is a small country. It is not rich in natural resources, gas or oil. We are not deeply religious. Therefore, people are our main resource, and education is our religion. In this connection, information technologies have become one of our main lines of development. The EU's IT Agency has resided in Estonia; the Estonians invented Skype; WIFI in Estonia may be compared to a human right which has to be provided always and everywhere. In our country the Internet has become an environment for business, public service, banking and other transactions. Facebook in the Estonian language is very popular and has become a communication channel for the youth. However, the Estonian language, just like other minority languages, still faces the challenge of gaining a victory over the English-speaking Internet. Or, rather, the Estonian language must enjoy equal status in the Estonian-speaking digital world along with the languages of international communication.

Among the residents of Estonia there are peoples whose native languages are great enough to enjoy their own digital environments, and the representatives of these peoples may use an opportunity of communicating on the Internet in their native languages. Thus, for the minorities of Estonia the invasion of the digital world is not a big danger, except for Ingrians, Mari, Komi and representatives of other minor Finno-Ugric peoples. In most cases, in view of the national composition, the most used is the Russian-speaking Internet environment, and the challenge here is attracting the Russian-speaking audience to use Estonian along with Russian for information search and communication in the Estonian Internet environment. No doubt, it is also necessary to develop Estonia's Russian-speaking information environment, but this is rather an issue of the national identity than of the Estonian language preservation.

Languages have always been in constant development and undergone natural changes. This is inevitable and even not bad at all – the entire world is in progress! At the same time it is essential, even more so for minority languages, to preserve a pure literary language by developing it incessantly through neologization. The Internet and its communication facilities are not necessarily the best options for preserving language purity, and this won't

happen by itself anyway. This issue should be made a subject of correct interaction in the communication environment (which is mostly the task for schools) and close consideration in the mass media information environment. There are minority languages on the verge of extinction which may serve as a bad example of disregarded neologization: when new notions appeared in the society (for instance, in the technological area or in the digital world), they chose the path of least resistance, i.e. they borrowed international words. As a result, the day may come when people realize that there is no use to produce new words in their language because, starting from a certain educational level, it lacks adequate vocabulary. In this respect, the Estonian language is out of danger because neologization has been the subject of our constant attention: we have established special neologization commissions in various areas (phytology, medicine, etc.). They meet and work on a regular basis, and we believe this practice must be continued. Also, you don't have to travel far for the examples of international words and incorrect forms of the native language that have been imported to the youth lexicon through social networks – the language which has been endlessly adapting to achieve higher communication convenience. All this undoubtedly impacts, and not always positively, the development of languages.

In conclusion I would like to note, taking Estonia as an example, that digital environment may maintain the development of the language and identity only until there are people speaking this language, and the number of these people should be sufficient enough to create an interesting and attractive Internet environment capable of meeting modern requirements. At the same time, it is insufficient to create a Facebook space in any language. It is essential to develop the vocabulary and keep an eye on the purity of the literary language, attaching much importance to it. As for minority languages that lack this attitude, they are sure to suffer from the impact of the world, which is moving deeper and deeper into the Internet, and may face extinction.

CONFERENCE FINAL DOCUMENT

Yakutsk Declaration on Linguistic and Cultural Diversity in Cyberspace

Preamble

The participants of the Third International Conference "Linguistic and Cultural Diversity in Cyberspace" are conscious that:

An overwhelming majority of peoples in the world have no statehood and sovereignty. As a rule, their languages are not the state languages of their country of residence because a majority of countries are multi-ethnic and multilingual. Even in the most beneficial situation, when the government and the larger, dominant ethnic entities display the utmost care of ethnic and linguistic minorities, most of languages are marginalized to varying extents. They develop or decline in the shadow of a major and fitter language dominant in its country and used in all spheres – political, economic, educational, cultural, scientific, etc.

The relationship between language and culture is one of interdependency. No language can develop outside the culture of the ethnos that created and speaks it. Culture is a function of language, and language is a vector for culture; no one exists without the other. Every time we talk about culture we are addressing language, and every time we look upon language we are reaching culture through it.

The linguistic and cultural diversity of humankind is the tip of an "iceberg" that includes cultural identities, the sense of belonging to a community, personal rootedness, intangible heritage, popular life-crucial knowledge and achievements over the centuries, proper interpretation of local content and much more.

The dissemination of multilingual information on the history, languages and culture of different nations enriches our ability to analyse facts, events and behaviours thanks to multiple viewpoints contributing to the promotion of tolerance and mutual understanding and a peaceful sustainable development of contemporary civilization. Cultural diversity and multilingualism are enablers of wellbeing and of the successful flourishing of humans.

Languages are stores of a rich and vast amount of human heritage and life-crucial knowledge, i.e. the knowledge necessary for health, well-being, and

participation in the local and worldwide community and economy, as well as necessary instruments for social life, the expression and dissemination of social and cultural traditions, self-identification and preservation of human dignity of their speakers, whether these are native to the territory or migrants.

Urbanization and globalization promote the assimilation of ethnic cultures and challenge their majority status, moving them ever farther into the margins. Knowledge and historical and cultural experience stored by these cultures vanish gradually, and the potential of those cultures and languages is reduced. Cultural and linguistic marginalization is an interrelated and interdependent process. A unique culture vanishes with the death of its languages. Meanwhile forecasts say and UNESCO has been continuously warning that more than half of the currently alive approximately 7,000 languages may become extinct within several generations.

Migration and human mobility have experienced an unprecedented increase over the past few years. For the first time ever in 2010, the majority of the world's population was predominantly urban, and this proportion continues to grow. By 2050, more than 70% of the world's population will live in a city. The high concentration of people in urban settings has resulted in an increasing linguistic diversity, a trend that will continue over the next years.

An increasing number of studies show that a well-planned strategy on managing diversity, including the languages of migrants, can lead to social and economic benefits for the society as a whole. Diversity can also create benefits as it increases the variety of goods, services and skills available in urban environments. The increased level of competences provided by diversity can also foster creativity, innovation and economic growth.

The role of ICTs

The global information society is forming rapidly, and the recent relevant social impact of improvements in Information and Communication Technologies (ICTs) makes this a turning point for the preservation of cultural and linguistic diversity.

Ever more people become active users of ICT, particularly of the Internet. It has become an inalienable and essential part of the life of the young majority, due to the extensive opportunities for communication, access to information and knowledge, self-expression, education, leisure and a greatly extended picture of the world. However, Internet services and information are mainly available in the dominant languages and the current absence of certain languages in cyberspace contributes to the widening of the already existing digital

information gap. While globalization encourages the merging of cultures and languages into a de facto standard, at the same time emerging ICTs potentially enable the exploitation of different languages and cultures and the flourishing of new alphabets and writing systems.

One of the main issues that is against the strengthening and diffusion of indigenous languages is that they are languages with no written tradition. In the context of our modern world, only a written tradition allows a language to become recognizable and useful. Cyberspace offers a unique opportunity for creating a writing tradition at low cost and with maximal possibilities of diffusion.

The digital era in which we live nowadays provides a unique opportunity to support the active promotion of well being, inter alia via the enabling of multilingualism and preservation of cultural diversity in cyberspace. Existing ICTs offer new opportunities to promote linguistic and cultural heritage for equal and universal access to life-crucial knowledge.

Being aware, in addition, that the increasing number of social media applications available and accessible may offer a relevant contribution to minority languages and cultural preservation and promotion thanks to services, communities and crowd initiatives, we must support initiatives to encourage them further.

Of course "availability" does not always mean "real use"; free or affordable access to ICTs is still a problem in many areas of the world and for many potential users. However, thanks to smart and mobile solutions both the new generation, and even those previously digitally excluded, have entered the digital age. Furthermore we are aware of the potential opportunities and threats connected to this process.

Guaranteeing linguistic and cultural diversity and the survival of every language and culture must be a common goal for humanity. Institutions ought to make commitments and assume responsibilities in relation to linguistic diversity and coexistence between languages.

Differences of languages and cultures should not create either manifest or hidden artificial obstacles for reasonable and fruitful cooperation among nations. This cooperation should be based on the equal treatment of all parties and it should not be governed by any cultural or linguistic prejudices.

The protection of linguistic and cultural diversity should include promoting and sustaining it in cyberspace, by enforcing digital opportunities for all languages.

Taking into account all the above and recalling the

- Outcomes of the First and the Second International Conferences on Linguistic and Cultural Diversity in Cyberspace (Yakutsk, Russian Federation, 2008 and 2011) as well as the Bamako International Forum on Multilingualism (Bamako, Mali, 2009), the 1st International Symposium on Multilingualism in Cyberspace (SIMC I - Barcelona, Spain 2009), the 2nd International Symposium on Multilingualism in Cyberspace (SIMC II – Brasilia, Brasil 2011) and the 3rd International Symposium on Multilingualism in Cyberspace (SIMC III – Paris, France, 2012);

- Universal Declaration on Cultural Diversity which says that "cultural diversity as a source of exchange, innovation and creativity is just as indispensable for humanity as biological diversity for Nature, and is a treasure shared by the entire human race";

- UNESCO Recommendation concerning the Promotion and Use of Multilingualism and Universal Access to Cyberspace;

- Key documents of the World Summit on the Information Society (Geneva, 2003, and Tunis, 2005) and the Vision for WSIS Beyond 2015 (WSIS+10, Geneva, 2014) which emphasize the importance of the preservation of cultural and linguistic diversity and suggest a set of measures necessary to achieve this goal;

The Conference agrees on the following recommendations:

General framework

Local, regional and central governments should play a wider and significant role in preserving, developing and sustaining local indigenous languages and their cultures by providing, in particular, resources for the development of digital tools, and the promotion of education and literacy for these languages to be present in cyberspace.

Governmental bodies at all levels should build professional communities and further develop necessary resources and tools (training courses, higher education courses, educational curricula, preparation of teachers, training for trainers, seminars, research, etc.) to strengthen linguistic and cultural diversity in cyberspace as digital language vitality is only sustainable where the language is spoken and taught. "Digital natives" should not be left alone in the digital arena. They belong to the always-on global communication community. They

create their own forms of language to exchange text messages, instant and multimedia messages as an additional legacy for future generations.

This can be done with the help of academia and thinktanks providing the expertise and identifying the best practices on language policies that not only promote all languages and put them on equal footing but also foster dialogue between co-existing languages in specific territories.

The development of natural language processing technologies (Text understanding, Question answering, Information querying, Speech Recognition, Speech Synthesis, (Spoken) Machine Translation and others) is a crucial step for ensuring equal digital opportunities for all languages. Mobile phone, Internet chat and social media interaction should be included in any definition of digital vitality and use of minority languages.

All institutions, organizations, associations and even individuals involved in language development, language planning or language promotion should set up collaborative projects to support lesser spoken languages, with a special attention to

- development of culture-based terminologies;
- production and dissemination of written materials and digital documents in these languages.

Local digital communities (not just read-only material) are also needed for digital ascent. Micro-grants to small communities (literary, theatrical, whatever brings people together) should be provided to document in their native language what they are doing.

Further examination should be made of the:

- identification of all the factors necessary to sustain linguistic diversity;
- link between digital vitality and spoken language vitality;
- impact of digital stillness on a spoken language.

The setting-up of language policies should include the state language, the national/regional language, and also migrant languages, as this is the only way to ensure that all citizens, regardless of their legal status in a territory, can be recognized equally by recognizing the plurality of their languages.

Diversity management of the communities enables a real representation of language minorities to promote stable and sustainable development.

Corpora, the lifeblood of modern computational linguistics, must be unencumbered by Copyright. A research exemption must be enshrined in the

legal framework. In addition national projects need to make their corpora not just searchable but also downloadable by ROAMing (randomize, omit, anonymize, mix). Open access to materials collected, as a precondition of funding, must be assured.

The significant part in linguistic diversity is linguistic diversity in scientific production. "Grey" literature naturally incorporates a greater linguistic diversity in scientific production particularly in developing and less developed countries. Promoting "grey" scientific literature can ensure linguistic diversity in knowledge production. Efforts should be deployed to raise the interest of the Academy in "grey" literature and acknowledge the alternative knowledge production it represents.

The open access movement makes scientific publications openly accessible and proposes that they give a greater priority to linguistic diversity so that the production of knowledge in multiple languages is promoted and facilitated, especially life-crucial knowledge.

All publicly funded translations of works should be available under free license for everyone to use and re-use without additional restrictions or exemption from copyright. This should apply both to works that are in the public domain as well as works still being under copyright, so they can be used as soon as copyright to the original work expires. International and national Copyright law systems should be amended in a way to allow all educational, scientific and non-commercial use of global cultural heritage.

In order to promote a widespread positive adoption of the notion of linguistic diversity and a favourable cultural climate that is the precondition for its flourishing, educational and informative tools need to be developed, aiming at correctly informing society at large and at educating the new generations about the value of multilingualism.

Efforts should be made to:

- support, enable and assist the documentation, protection and promotion of regional, minority and endangered languages;
- bring together specialists in specific areas of the world with members of endangered language communities in these areas.

It is very important to ensure long-term preservation of audio and video recordings reflecting linguistic and cultural diversity, specifically of orally transmitted cultures, key elements to ensure demo-anthropological studies. Over the past 60 years, rich collections of such audio-visual recordings have been produced, forming the basis of our present knowledge. Nowadays they

can only be preserved by digitization and proper digital preservation of data. Presently, however, the availability of replay equipment for magnetic audio and videotapes is dramatically shrinking which may bring digitization programmes to a halt. Adequate measures must be taken to respond to this threat in order to prevent such unprecedented loss of irreplaceable documents of the linguistic and cultural diversity of humankind.

At the political level

1. All stakeholders should seek to facilitate the emergence of knowledge societies respecting human rights and values and based on four principles: Promoting freedom of expression in traditional and new forms of media, including the Internet; Access to quality education for all; Respect of cultural and linguistic diversity; and Universal access to information and knowledge, especially in the public domain;

2. UNESCO, especially through its Information for All Programme (IFAP), should pursue in cooperation with relevant UN agencies, IGOs and NGOs the efforts for further development and promotion of ethical, legal and societal principles and norms for preserving linguistic and cultural diversity worldwide, in particular in cyberspace;

3. All stakeholders should encourage governments to enact and implement more effective national policies in support of preservation of linguistic and cultural diversity;

4. All the stakeholders should continue to promote and support the creation and free dissemination of language resources (alphabets, diacritic marks, phonetic language resources, Wikipedia, Wiktionaries, and related technical means (e.g. spell checkers, and generally speaking natural language processors), with specific reference to the use of virtual keyboards), necessary for the use of indigenous and minoritized languages in cyberspace ensuring equal digital opportunities for all languages;

5. UNESCO and its Member States should continue to develop with the relevant IGOs and NGOs policies to enhance the presence (localization and content) of all languages in cyberspace, based on media and information literacy, access to resources and promotion of participation, developing programmes of inclusion of knowledge from languages unrepresented on the Internet, creating a comprehensive and sustainable set of indicators, and promoting a comprehensive view of the digital divide which encompasses the content and linguistic divide;

6. Governments should strengthen existing discussion platforms involving all concerned stakeholders in a continuous interdisciplinary debate on preservation of linguistic diversity in the context of current socio-cultural transformations in globalizing information society/knowledge societies;

7. All stakeholders, especially in academia should develop interdisciplinary research and comprehensive study on the various political, social and cultural challenges especially in education with regard to the preservation of linguistic diversity;

8. All stakeholders, especially governments, scholars and experts in academia, should develop and strengthen educational and awareness-raising programmes, especially among the youth, to form stronger respect of linguistic and cultural diversity and deeper understanding of the necessity to preserve all languages particularly minority/minoritized ones.

The Conference also especially recommends UNESCO, and other relevant international, regional and national stakeholders, to

a) develop and propose an international act on global linguistic and cultural diversity in cyberspace as well as in other communication spaces;

b) stimulate the creation of a worldwide network of competence centres for the study and promotion of multilingualism in cyberspace and for sharing expertise on the subject;

c) set up a working group representing each continent to identify language policies, initiatives, and digital opportunities that best respect the linguistic reality of the territory;

d) invite the National IFAP committees to support the development of culture-based terminologies and promote online publication of indigenous and minority history, events, raw data (Newspapers, Books, Radio-TV Broadcast, Videos, etc.), as well as the production and distribution of annotated corpora, lexica, dictionaries, grapheme-phoneme convertors, parallel corpora, etc. allowing for the development of language technologies;

e) invite the National IFAP committees to create and support a specific action aimed at activating crowdsourcing in order to address autoch-

thonous and minoritized language content digitization (Adopt an endangered language/culture!) and constitute related communities;

f) emphasize the role of the broader literacy context for digital language ascent (work the chain of literacy);

g) conduct comparative study on national policies to sustain languages, including linguistic preservation and development in cyberspace;

h) conduct a survey of the state-of-the-art regarding language resources and technologies for all languages;

i) set up a clearing-house on language technologies to be used for identifying and assessing digital language development;

j) start focusing (including through IFAP) on the possibility of using cyberspace as a laboratory for creating writing traditions for endangered and minority languages;

k) further work on the development of language technologies for minority languages, which can be easily adapted for application from one language to another (easier localization techniques), especially within language families;

l) ensure the availability of language technologies for the largest number of languages through the cooperation of all stakeholders (Member States and regions, public and private research laboratories, industries) under the general coordination of UNESCO;

m) design an Index of Digital Language Diversity, as an instrument for measuring the digital language diversity of a given region and for assessing the type of intervention needed to ensure all languages equal digital opportunity;

n) invite universities in all countries to submit project proposals to international institutions (such as UNESCO Chairs and UNITWIN) to promote multilingualism and cultural diversity in cyberspace;

o) support the documentation of endangered languages in their area and provide an inventory of existing material;

p) start working on the modernisation of the existing Atlas of World Languages by using advantages of modern ICTs;

q) set up fund-raising events and look for possible sponsors of the work in the area.

A Virtual Observatory for Multilingualism and Digital Language Diversity (possibly as a part of the IFAP Information Society Observatory) should be set up in order to:

1. accumulate reliable and up-to-date data about the presence of the world's languages on the Internet, the availability of digital technologies supporting languages, and of infrastructural conditions enabling presence of languages in the digital world;

2. accumulate publications on academic and civil society projects concerning multilingualism in digital world;

3. closely monitor existing language policies that best respect the linguistic reality of the territories, and the efforts and initiatives taking place worldwide to support multilingualism both offline and online;

4. help identify best practices being developed in each continent;

5. highlight what works best in each geographical location and to monitor the extent to which language diversity is digitally reflected;

6. map the researchers working on the subject and the fields of knowledge.

The Conference finally urges UNESCO and other relevant international, regional and national stakeholders to initiate a preparatory process for a World Summit on Multilingualism as it is highly desirable for the preservation and development of the world's languages and cultures in the epoch of galloping globalization.

<p align="center">*********</p>

This document was produced through a collaborative process involving participants from the following countries – Albania, Argentina, Austria, Azerbaijan, Belarus, Botswana, Brazil, Bulgaria, Central African Republic, China, Colombia, Czech Republic, Dominican Republic, Ecuador, Estonia, Finland, France, Hungary, India, Israel, Italy, Japan, Kazakhstan, Kyrgyzstan, Latvia, Macedonia, Moldova, Netherlands, Nigeria, Oman, Peru, Poland, Republic of Korea, Republic of Maldives, Russian Federation, Rwanda, Slovakia, Spain, Sri Lanka, Sudan, Sweden, Syria, Thailand, Togo, Turkey, UK, USA.