ISCA SALTMIL SIG:

Speech and Language Technology for Minority Languages

Briony Williams[1], Kepa Sarasola[2], Climent Nadeu[3], Donncha Ó'Cróinín[4], and Bojan Petek[5]

[1]

briony@swinnerton.demon.co.uk

[2]Faculty of Computer Science
University of the Basque Country
jipsagak@si.ehu.es

[3]Linguistics Institute of Ireland
ITE
doc@ite.ie

[4]TALP Research Centre
UPC
climent@talp.upc.es

[5]Interactive Systems Laboratory
University of Ljubljana
Bojan.Petek@Uni-Lj.si

Abstract

This paper presents International Speech Communication Association (ISCA) Special Interest Group (SIG, http://www.isca-speech.org/sig.html) on Speech And Language Technology for MInority Languages (SALTMIL). The paper gives an overview of the group's mission and discusses its past and present activities.

1. Introduction

The start of a new ISCA SIG on Speech And Language Technology for MInority Languages (SALTMIL) was announced in October 1999 [1,2]. Founding committee members were Dr Briony Willliams, formerly with CSTR, Edinburgh University, who became the SALTMIL Chair and Liaison representative, Professor Climent Nadeu, Universitat Politecnica de Catalunya, who took responsibilities of SIG Secretary and Dr Donncha Ó Cróinín, Linguistics Institute of Ireland who acted as the SIG Publicity Officer.

Initial activities of the SALTMIL SIG were aimed to establish an Email discussion list (now at [3]) and the web site (now at [4]).

2. SALTMIL Issues, Aims and Objectives

Local language could be defined as the human language non-visible in information system mediated natural interactivity of the information age. The local languages frequently cope with difficulties in appropriate amount of language resources in standardized electronic form, expertise and/or funding to overcome technological neglect in the future [5].

On the other hand, several recent research programs for global languages focused on the next generation of conversational interfaces. Their fundamental goal is to create speech enabled multimodal systems that scale gracefully across modalities. Such interfaces include speech, graphics, gesture and support complex conversational interaction principles comparable to the human-human interactivity. The systems typically integrate spoken language dialogue systems, including multimodal communication and web-based data handling tools. The long term goal of this research is therefore to transform the present computer systems to become transparent in communication task and to support similar communication patterns as experienced in common interpersonal communication [6, 7].

Given these premises the general aims of the SALTMIL SIG represent promotion of research, development and education in the area of Human Language Technology (HLT) for less prevalent (lesser-used, lesser-studied, minority, regional or local) languages. Main group activities include organization of workshops/conferences with the focus on issues pertaining to the less prevalent languages. Another important issues are dissemination of information between researchers in less prevalent and major language groups as well as provision of guidelines that should initiate and help to promote common projects between all researchers in the HLT area [5].

3. Past activities

Briony Williams proposed two notable workshops that were successfully organized so far. The first one was entitled the Language Resources for European Minority Languages, and was held in May 1998 in conjunction with the First International Language Resources and Evaluation Conference (LREC 1998) in Granada, Spain [8]. The next workshop was organized in May 2000, on the topic Developing Language Resources for Minority Languages: Re-useability and Strategic Priorities, just before the Second International Language Resources and Evaluation Conference (LREC 2000) in Athens, Greece [9].

Both workshops succeeded as lively forums for exchange of information among researchers and brought unique opportunities that helped to shape the focus of forthcoming SALTMIL activities.

3.1. The LREC 1998 workshop

The first workshop that also included the SALTMIL SIG proposal was organized on Wednesday May 27, 1998. About 35 people attended the workshop. Five oral papers and 12 poster papers were presented, all available online at [8].

Further details on the workshop are detailed at [8]. Dr Nicholas Ostler, President of Foundation for Endangered Languages, wrote an interesting online overview paper accessible at [9].

3.2. The LREC 2000 workshop

One day workshop on the topic Developing language resources for minority languages: reuseability and strategic priorities was held on the afternoon of May 30, 2000 in Athens, Greece. The aim of the workshop was to bring together researchers developing language resources for minority languages in order to build contacts and share experience [10].

The minority or lesser used languages of the world (e.g. Basque, Welsh, Breton) are under increasing pressure from the major languages. Some of them (e.g. Gaelic) are endangered, but others (e.g. Catalan) are in a stronger position. However, the situation with regard to language resources is fragmented and disorganized. Some minority languages have been adequately researched linguistically, but most have not, and the vast majority do not yet possess basic speech and language resources (such as text and speech corpora) which are sufficient to permit commercial development of products [10].

If this situation were to continue, the minority languages of the world would fall a long way behind the major languages, as regards the availability of commercial speech and language products. This in turn will accelerate the decline of those languages that are already struggling to survive, as speakers are forced to use the majority language for interaction with these products. To break this vicious circle, it is important to encourage the development of basic language resources [10].

The workshop was a step towards encouraging the development of such resources. The aim was to disseminate information on existing projects and possible future strategies, as well as to form personal contacts and share best practice. This will make it easier for isolated researchers with little funding and no pre-existing resources to begin developing language resources that are maximally useful [10].

About 39 people registered for the event. The workshop program included four oral papers and 14 poster presentations, detailed at [10], including the first meeting of the ISCA SALTMIL SIG, described in [11,12].

3.3. The Multilinguae Congress

<Kepa: please report on the content of http://www.gaia.es/multilinguae/htm/ihtm/program.htm and on action points taken at our informal SALTMIL meeting there>

4. Present activities

<Donncha: could you please report on progress regarding the agreed action points taken in Donostia>

5. Discussion

<Briony's and Climent's slot>

6. References

URL list was accessed in March 2001:
[1] http://www.isca-speech.org/escapads/iscapad18
[2] http://www.emich.edu/~linguist/issues/10/10-1542.html
[3] http://groups.yahoo.com/group/saltmil/
[4] http://isl.ntftex.uni-lj.si/SALTMIL/
[5] http://isl.ntftex.uni-lj.si/SALTMIL/aims.html
[6] http://www.hltcentral.org/projects/
[7] http://nespole.itc.it/
[8] http://isl.ntftex.uni-lj.si/SALTMIL/lrec98.html
[9] http://isl.ntftex.uni-lj.si/SALTMIL/review.html
[10] http://isl.ntftex.uni-lj.si/SALTMIL/lrec00.html
[11] http://isl.ntftex.uni-lj.si/SALTMIL/athmins00r.html
[12] Petek, B., "Developing Language Resources for Minority Languages: Reusability and Strategic Priorities", *ELRA Newsletter*, 20-21, July-September 2000. (ISSN 1026-8200)

GAIA, the Telecommunications Cluster Association of the Basque Country (http://www.gaia.es), supported by the European Commission and with collaboration from Eusko Ikaskuntza (http://suse00.su.ehu.es) and the IXA Group (http://ixa.si.ehu.es), organised the II International Multimedia and Minority Languages Congress (http://www.gaia.es/multilinguae), which was held on the 8th and 9th of November in the Miramar Palace in Donostia-San Sebastián.

We know by intuition that for a community such as the Basque Country (with a longstanding history, bilingual and bicultural wealth, with a desire for future projection in the Information Society, etc.), having a focal reference point for these experiences would be a highly important competitive and strategic advantage. Also, the consolidation of a line of studies in this field would be a reference point in Europe and internationally.

Over 80 professionals from the world of technology, GAIA associates and collaborators, and specialists in minority language studies in Europe came together at this Congress, where European minority languages played a leading role.

Basque (spoken by some 500,000 people in the world), Gaelic (spoken in Ireland, Scotland, Brittany and Wales), Mennonite (spoken in Germany, Canada and the United States), Cimbro (spoken in Italy and considered to be the smallest minority language today in Europe, with less than 500 speakers) and Catalan (although spoken by approximately 10 million people all over the world), amongst others.

Among the large number of speakers who took part in this Congress were the Director-General of the Society for Basque Studies, José María Vélez de Mendizabal; Joserra Azurmendi, from ADUR Software Productions, S.Coop. (an outstanding Basque company dedicated to Basque (http://www.adur.es) software development); and Xabier Artola, from IXA (experts in natural language processing).

Other participants were Caroline Philips from Philips Consulting (a company which develops electronic commerce supports in Basque); and Alastair MacPhail

(http://www.europa.eu.int/comm/education/callg.html), a European Commission official who provided information on financing for the development of future projects. Also participating in the convention were other specialists in minority language studies at an international level such as Donncha O'Croinin (Ireland), Sara Scardoni (Italy), Peter Wiens (Germany), Jon Patrick (Australia) and Bojan Patek (Slovenia).

GAIA's general aim with regard to this Convention was to promote and boost the capacities of the new technologies, especially the Internet, in the development and preservation of minority languages, and to foment what is known as the Language Industry in the Basque Country.
As regards research, the study of virtual communities of linguistic minorities began to develop, especially in the USA, at the beginning of the 1990s with a programme of the Department for Indian Affairs on "Communication Technologies and the Building of Native Communities in the USA". In solidarity with this programme various Latin-American and European minority communities began to promote programmes on virtual communities.

At a European level some programmes and projects are being implemented, giving rise to reference studies. As an example we can mention the "Minority language community empowerment and digital democracy" project of 1998, headed by the University of Wales, with the participation of Welsh, Irish and Finnish linguistic minorities; and the ACCORDE project, which explored in depth and using historical cases the use of advanced communications in regional development and cohesion.

The most advanced European country as regards this issue is Sweden, which has carried out various research projects on the linguistic minorities within its territory and on the impact of Information Technologies on the linguistic education of these minorities.

Anthropologists such as Bonnie Nardi and Alice McInnes are studying the change taking place in the indigenous communities and linguistic minorities in the USA as a result of the introduction of the new information technologies.

GAIA, as a Telecommunications Cluster Association, aspires to foster each and every one of the activities of interest to be developed by the ICT industry in the Basque Country.

We therefore hope that the development of this kind of activities will encourage the advancement and use of the minority languages with the aid of the New Technologies.

Some reference web pages on minority languages:

http://www.gaia.es/multilinguae

http://www.eblul.org/

http://www.europa.eu.int/comm/education/callg.html

http://www.volkswagen-stiftung.de/infotext/infodoku.htm

http://www.solki.jyu.fi/elc/elc.htm