

Developing language resources for minority languages: re-useability and strategic priorities

Pre-Conference Workshop Announcement and Call for Participation

Held in conjunction with the

**Second International Conference on Language Resources and Evaluation
(LREC 2000) Athens, Greece**

Tuesday, 30th May 2000 (14:30 - 20:00)

1. Workshop announcement
2. Workshop scope, aims and program
3. Organising committee
4. Important dates
5. Contact person for the workshop
6. Submissions
7. Workshop registration
8. Conference information

1. Workshop announcement

There will be a one-day workshop on the theme of "Developing language resources for minority languages: re-useability and strategic priorities" on the afternoon of May 30th

2000 at Athens, Greece (preceding the 2nd International Language Resources and Evaluation Conference, LREC2000). The aim of the workshop is to bring together those who are developing language resources for minority languages, in order to build contacts and share experience. The workshop will include the first meeting of the ISCA SALTMIL SIG: "Speech and Language Technology for Minority Languages".

2. Workshop scope, aims and program

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 disorganised. 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.

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.

The workshop is a small step towards encouraging the development of such resources. The aim is to disseminate information on existing projects and possible future strategies, as well as forming personal contacts and sharing 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.

The workshop will also incorporate the first meeting of the ISCA "SALTMIL" SIG: "Speech and Language Technology for Minority Languages".

The Workshop Program

- 14:30

Registration, and preparation of posters

- 14:50

Welcome Bojan
Petek, University of Ljubljana, Slovenia

- 15:00

Strategic priorities for the development of language technology in minority
languages
Kepa Sarasola, University of the Basque Country, Spain

- 15:20

Language engineering resources for minority languages
Harold Somers, University of Manchester Institute of Science and Technology,
UK

- 15:40

Linguistic Exploration: New Methods for Creating, Exploring and Disseminating
Linguistic Field Data
Steven Bird, Linguistic Data Consortium, USA

- 16:00

Funding for research into human language technologies for less prevalent languages
Bojan Petek, University of Ljubljana, Slovenia

- 16:20

Coffee

- 16:40

General discussion of talks

- 17:15

Poster session

- 18:45

SALTMIL SIG first meeting

- 20:00

Finish

The contributed poster papers will focus on existing projects in the field, with the opportunity to share useful information. This includes (but is not limited to) topics such as:

- Presentations of existing speech and text databases for minority languages, with particular emphasis on software tools that have been found useful in their development.
- Presentation of existing lexicons for minority languages, with particular emphasis on fast production methods.

The scheduled poster session includes the following presentations:

The Development of Language Resources for Maltese

Paul Micallef*, Mike Rosner** (*Dept. Communication and Computer Engineering, University of Malta)

Courseware Based on Speech Technology for Breton Language Pronunciation Learning: S

Guy Mercier*, Jacques Siroux*, Francis Favereau**, François Louis*** (*ENSSAT, CORDIA

Neutral Lingware: A Need for Technologically Weak Languages

Lluís de Yzaguirre i Maura (Institut de Lingüística Aplicada, Universitat Pompeu Fabra, Barcelona)

The Spanish-Galician and Galician-Spanish MT System: How	to re-use the existing Galician
Inés Diz Gamallo*, Liliana Martínez Calvo** (*Centro Ramón	Pineiro para a Investigación e
Electronic Dictionary of Pronunciation and Usage of the Graecanic	Dialect of Southern Italy
George Kokkinakis, Helen Coutsogeorgopoulos, Evangelos Dermatas,	George Kaitsas (W
Unit Selection based on Diphones for Catalan Text-to-Speech	Conversion
Roger Guaus i Térmens, Ignasi Iriondo Sanz (Secció Tecnologies	de la Parla, Departamen
The State of the Art of French Creole Language Resource Engineering	
Marilyn Mason (Mason Integrated Technologies Ltd, Boston,	Massachusetts, USA)
Segre: An Automatic Tool for Grapheme-to-Allophone Transcription	in Catalan
P. Paches*, C. de la Mota**, M. Riera**, M. P. Perea***,	A. Febrer*, M. Estruch**, J. M. Ga

Tools

and Basque Language Databases Developed in the Aholab

Laboratory

Borja

Edgar Ríos, SuperNiles, Área (Dance, Electrical, Mechanical, Biología, Spain) Telecomunicaciones, Escuela

Digital

Sounded Lexicon of Nenets

Marina

a Ljublinskaja*, Tatiana Sherstinova**, Elena Kuznetsova*

(*Institute of Linguistic Res

A Corp

Corpus of Written Finnish Romani Texts

Lars B

Borin (Department of Linguistics, Uppsala University,

Uppsala, Sweden)

Comp

Outer Fund of the Tatar Language: Information Resources

of Vocabulary and Textual S

Kamil

R. Galiullin (Kazan State University, Kazan, Republic

of Tatarstan, Russia)

Czech

Speech Corpus for Development of Speech Recognition

Systems

Vlasta

a Radová, Josef Psutka, Lubos" S"mídl, Petr

Vopálka, Filip Jurcícek (University of We

Creating High-Quality, Large-Scale Bilingual Knowledge Bases

Using Minimal Resources

Davide Turcato, Fred Popowich, Paul McFetridge, Janine Toole

(Natural Language Lab, S

3. Organising committee

- [Briony Williams](#) , CSTR, University of Edinburgh, Scotland, UK
- [Donncha O'Croinin](#) , Linguistics Institute of Ireland/ITE, Dublin, Ireland
- [Climent Nadeu](#) , Universitet Politecnica de Catalunya, Barcelona, Spain
- [Kepa Sarasola](#) , University of the Basque Country, Spain
- [Bojan Petek](#) , University of Ljubljana, Slovenia

4. Important dates

- Deadline for workshop abstract submission

11 February 2000

- Notification of acceptance

3 March 2000

- Final version of paper for workshop proceedings

28 March 2000

- Workshop

30th May 2000 (afternoon)

5. Contact person for enquiries about the workshop

[Briony Williams](#)

Centre for Speech Technology Research
80 South Bridge
Edinburgh EH1 1HN
Scotland, UK
Tel: +44 131 650 2790
Fax: +44 131 650 6351

6. Submissions

slt4d09

Papers are invited that will describe existing speech and language resources for minority languages (speech databases, text databases, and lexicons), also papers based on the analysis of these resources. All contributed papers will be presented in poster form. Each submission should show: title; author(s); affiliation(s); and contact author's e-mail address, postal address, telephone and fax numbers. Abstracts (maximum 500 words, plain-text format) should be sent to:

Donncha O'Croinin, ITE, 31 Fitzwilliam Place, Dublin 2, Ireland.

email: doc@ite.ie

Those who wish to attend without offering a paper are asked to indicate their interest to Donncha O'Croinin (doc@ite.ie) in order to receive their own copy of the final programme and registration details.

The final version should not be longer than 4,000 words or 10 A4 pages. Instructions for formatting and presentation of the final version will be sent to authors upon notification of acceptance.

7. Workshop registration

The registration fee for the workshop is:

- If you are not attending LREC: 80 EURO
- If you are attending LREC: 40 EURO

The fee includes a coffee break and the workshop proceedings.

Participation in the workshop is limited by the venue. Requests for participation will be processed on a first come first served basis. Registration is handled by the [LREC Secretariat](#).

8. Conference information

General information on LREC2000: <http://www.icp.grenet.fr/ELRA/lrec2000.html>

Specific queries about the conference and registration for the workshop:

LREC Secretariat
ILSP
6, Artemidos & Epidavrou Str
15125 Marousi
Athens
Greece
email: LREC2000@ilsp.gr
Tel: +30 1 6800959
Fax: +30 1 6856794