



2013 annual activity report

Content

Word by the President & Secretary General of ELRA

2013 at a glance

ELRA Members

ELRA Activities and Projects

LREC, the Language Resources and Evaluation Conference

Word by the President & Secretary General of ELRA

The Board has been drawing a new strategy for ELRA for the next 5 years in order to consolidate ELRA as a key player in the HLT community, focusing on anticipating the community's new expectations as well as strengthening its relationship with the European Commission. New trends are perceived through many initiatives popping up within the community, especially within the framework of the EC-funded R&D programs (FP7, CIP, Horizon2020, CEF, etc.). In line with this strategic orientation, several actions are being put into operation:

- The establishment of the **International Standard Language Resource Number (ISLRN)** (<http://www.islrn.org>), a unique identifier to be assigned to each LR allowing a unique identification of a resource, regardless from its physical storage, its availability or the associated license.
- The organisation of the **NLP12 meeting**, with the 12 major NLP associations and institutions from Language Resources and Technologies, Computational Linguistics, Spoken Language Processing and Digital Humanities aiming at boosting their co-operation and strengthening the bridges between various communities.
- The exploitation of the **META-SHARE** platform in order to become ELRA's next instrument of sharing LRs, including the development of new extensions such as a **legal Wizard** to help the right-holders share and/or distribute their resources under the right license.
- The celebration of the association's **18th anniversary** in November 2013, in Paris, through the organisation of the workshop *ELRA International Workshop on Sharing Language Resources: Landscape and Trends*.

"We continue to vision 2013-2015 with emphasis on a renovation of ELRA policy and activities. ELRA continues to discuss potential extension of its membership to new types of members in particular to attract individuals so it can serve them better."

Nicoletta Calzolari, President



Khalid Choukri, Secretary General



ELRA Board

Board Officers

Nicoletta Calzolari (ILC-CNR, Italy), President

Nick Campbell (Trinity College, Ireland), Vice-President

Isabel Trancoso (INESC, Portugal), Vice-President

Maria Gavrilidou, (ILSP, Greece), Secretary

Andrejs Vasiljevs, (Tilde, Latvia), Treasurer

Honorary President

Joseph Mariani (IMMI & LIMSI, France)

Board Members

Thierry Declerck, (Language Technology Lab DFKI, Germany)

Marko Grobelnik, (J. Stefan Institute, Slovenia)

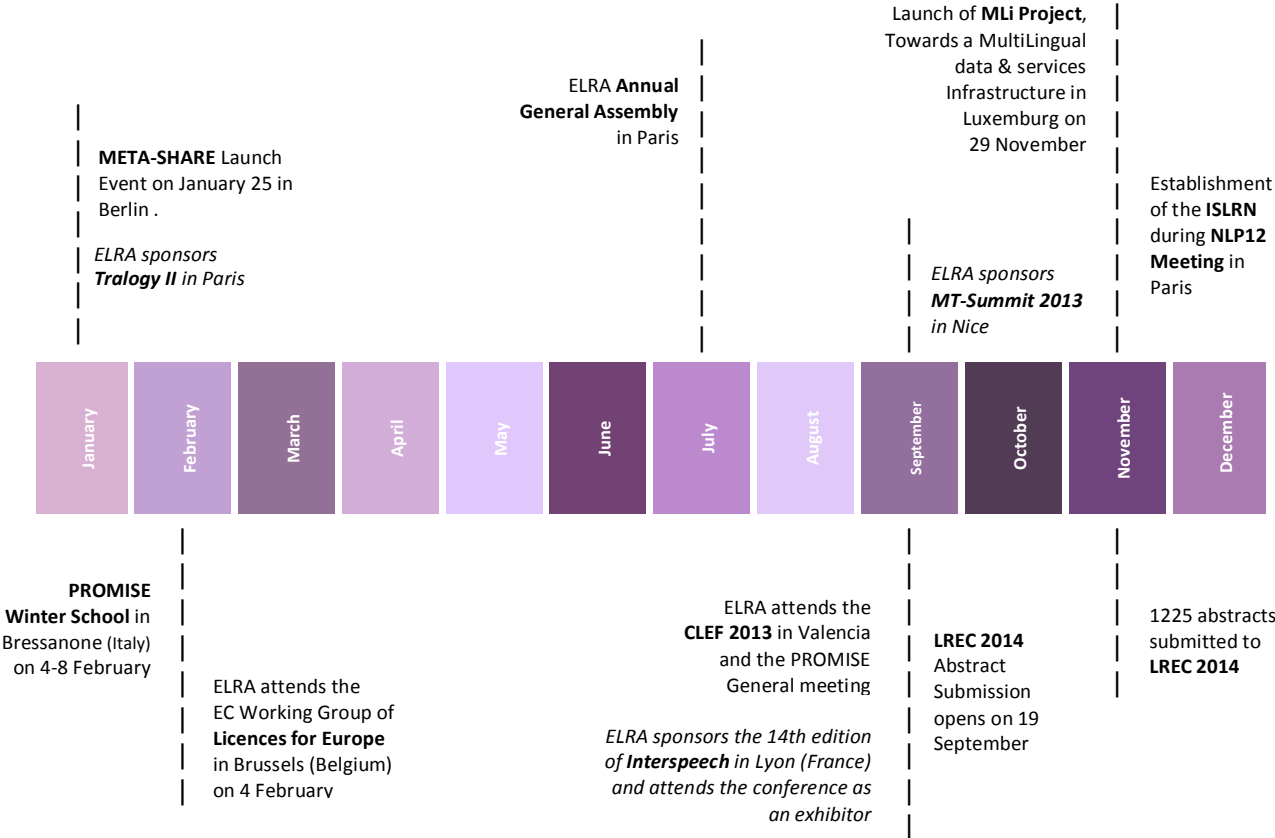
Henk van den Heuvel, (CLST, Radboud Universiteit, The Netherlands)

Frédérique Segond, (Objet Direct, France)

Secretary General

Khalid Choukri, (ELDA, France)

2013 at a glance



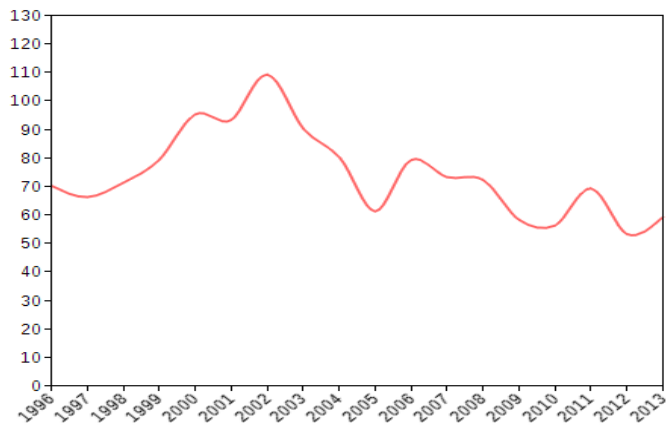
ELRA Members

ELRA membership scheme is **institution-based**. Any organisation, public or private, European or non-European, can join. However, full membership, with voting rights, is available only to organisations legally established in Europe, as per the article 5 of the Statutes.

2013 Membership at a glance

Who are our members?

59	total number of members
23	new members joined in 2012
Members having renewed their membership	36
7	members joined in 1995 and have renewed membership continuously ever since
European members	80%
61%	European non profit organisations



Membership Over the Years

Where are our 2013 members?



*The tags show the countries.

Services to members

- **Price discounts** on the resources, up to 70% reduction on some items, including the resources produced by ELRA.
- **Discounted fees** for members registering to the biennial conference LREC.
- **Legal and contractual assistance** are also provided to the members of the association
- **Regular information** through the monthly Members' News bulletin and the updates brought to the www.elra.info web site.
- Special prices are proposed to the members for the **Language Resources and Evaluation (LRE) Journal**, published by Springer.

Legal Support HelpDesk

Using, producing, sharing or distributing Language Resources can trigger legal questions related to Intellectual Property Rights (IPR) management which are not always easy to sort out. For nearly 20 years now, ELRA has established close co-operation with legal experts to clear such IPR issues, to design licensing schemas and draft licenses, but also to provide assistance on any contractual or legal matter that may arise during the Language Resources life cycle, including acquisition, production, sharing, or distribution phases.

With this Helpdesk service fully dedicated to IPR issues, we are extending our legal support to the whole Human Language Community. This Helpdesk provides services similar to those offered by the META-SHARE network.

Please contact us by sending a message to elra-ipr-helpdesk@elra.info.

Support HLT-related Events

A number of events have been sponsored by ELRA in order to support and endorse their activities within our field. In 2013, these are:

- TRALOGY II
- ERRARE 2013
- MT SUMMIT 2013
- Interspeech 2013

In addition, ELRA continues to support the Linguist List.

Our members are welcome to contact us if they seek such support for events (conferences, workshops) that focus on HLT and preferably on LRs and Evaluation.

LRE Friends

The Forum of Language Resources and Evaluation Friends (LRE-F) was established at the last LREC in Istanbul. The LRE Friends are individuals interested in Language Resources and Evaluation. They can be scientists, students or professors, involved in HLT research activities in universities, small and medium companies or international groups, decision-makers or project managers in large public institutions, etc.

ELRA Activities and Projects

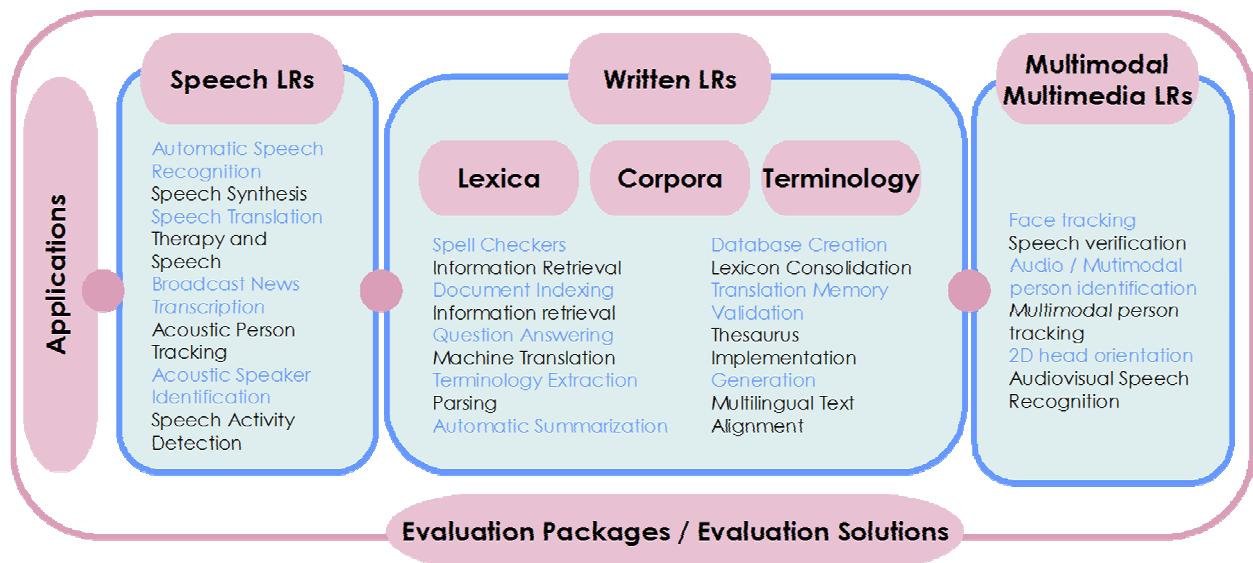
In addition to the promotion of language resources for the Human Language Technology (HLT) sector through market surveys, publication of a regular newsletter and organisation of LREC conference and other workshops, the activities of ELRA are organised around the following services:

- Identification, collection and distribution of language resources
- Production of language resources
- Validation of language resources
- Evaluation of systems, products, tools, etc., related to language resources
- Standardisation

The promotion of the language resources production also includes our support of both the infrastructure for evaluation campaigns and the development of a scientific field of language resources and evaluation, e.g. via the LREC conference.

Language Resources Catalogue

The Language Resources collected by ELRA are made available to the public through the ELRA Catalogue that can be accessed online at <http://catalog.elra.info>*. The distribution of resources, which cover a wide variety of languages and belong to different modalities, is shown below:



The term **Language Resources** refers to sets of language data and descriptions in machine readable form, used in many types of areas, components, systems, applications, including the creation and evaluation of natural language, speech and multimodal algorithms and systems, the software localisation and language services, the language-enabled information and communication services or the knowledge management; e-commerce, e-publishing, e-learning, e-government.

* Now also part of the META-SHARE network at www.meta-share.net

Universal Catalogue

The Universal Catalogue is an important identification feature. Information regarding Language Resources (LRs) identified all over the world are gathered in this publicly available repository. The LRs are generally located by the ELRA team, but external feedback from our members, collaborators, and web-site visitors is also included. Our aim is to provide researchers and developers information about existing LRs and spare them the effort of searching or rebuilding similar resources. The Universal Catalogue is a collaborative feature: through a simple form, anyone interested can add some basic information, point to an existing language resource or enrich the current description of a LR already present in the Universal Catalogue. Up to now, 1647 LRs have been listed and they distribute as follows:

- Tools: 57
- Speech corpora: 569
- Written corpora: 720
- Lexicon: 203
- Terminology: 24
- Multimodal/ Multimedia : 74

The Universal Catalogue is accessible through the following web site: <http://universal.elra.info/>.

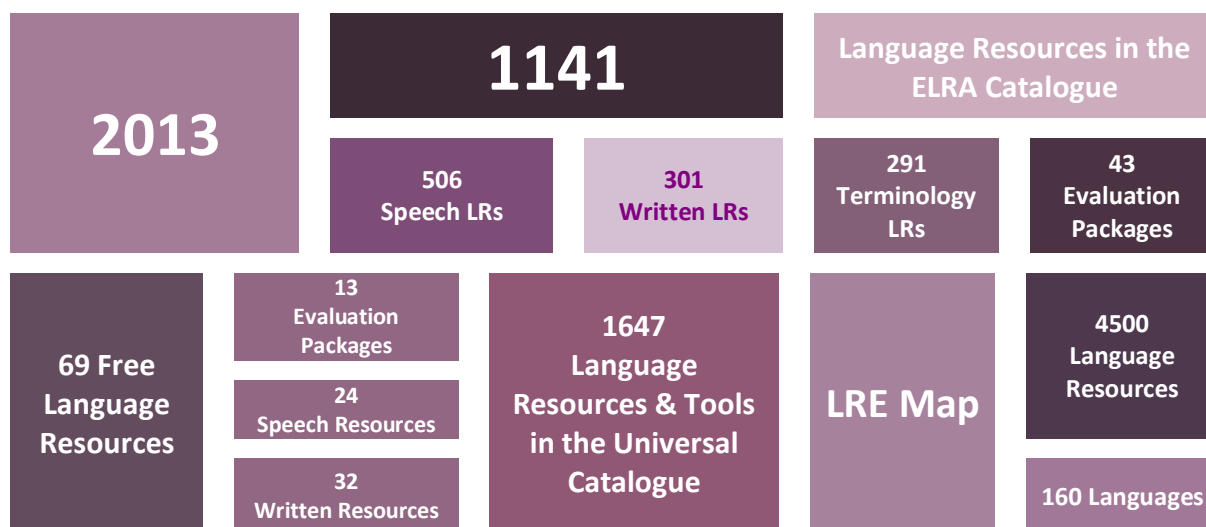
LRE Map

The initiative for LR & Tools identification, the LRE Map

Initiated by ELRA and FlareNet at LREC 2010, the LRE Map is a new mechanism intended to monitor the use and creation of language resources by collecting information on both existing and newly-created resources during the submission process. Nearly 2000 language resource forms have been filled in. Apart from providing a portrait of the resources used by the community, of their uses and usability, the LRE Map intends to be a measuring instrument for monitoring the field of language resources. The feature has been so successful that it has been implemented also at COLING 2010 and EMNLP 2010, while other major conferences have adopted it since then.

The LRE Map is now part of the LREC submission process which has been slightly updated for the LREC 2012 submission to consider the changes brought to the 2012 version of the Map. Further to the very high number of submissions to LREC 2012 (1027 for both oral and poster papers), the number of LR forms have increased substantially. Globally, 4260 LR forms and 3015 LR types have been filled in within all the conferences which have adopted the LRE Map.

ELRA Catalogues in the numbers



New LR in 2013 & Distribution

The Language Resources collected by ELRA are made available to the public through the ELRA Catalogue and in 2013, a number of new agreements have been signed with providers. Some are shown below:

Speech LRs

Speech Desktop/Microphone

GlobalPhone Resources

GlobalPhone Pronunciation Dictionary

- ELRA-S0340: French, ELRA-S0341: German
- ELRA-S0348: Japanese, ELRA-S0350: Arabic
- ELRA-S0351: Bulgarian, ELRA-S0352: Czech, ELRA-S0353: Hausa
- ELRA-S0354: Polish, ELRA-S0355: Portuguese (Brazilian)
- ELRA-S0356: Swedish, ELRA-S0358: Croatian
- ELRA-S0359: Russian, ELRA-S0360: Spanish (Latin American),
- ELRA-S0361: Turkish, ELRA-S0362: Vietnamese
- ELRA-S0363: Chinese-Mandarin , ELRA-S0364: Korean

Speech Telephone Resources

- ELRA-S0365 aGender

Broadcast Resources

- ELRA-S0349 Quaero Broadcast News Extended Named Entity corpus

Written LRs & Evaluation Packages

Written Corpora

- ELRA-W0057 PANACEA English-French and English-Greek parallel corpus acquired for Environment domain
- ELRA-W0058 PANACEA English-French and English-Greek parallel corpus acquired for Labour Legislation domain

PANACEA Resources

PANACEA Environment Monolingual Corpora

- ELRA-W0063: English , ELRA-W0065: French, ELRA-W0067: Greek
- ELRA-W0069: Italian, ELRA-W0071: Spanish

PANACEA Labour Monolingual Corpora

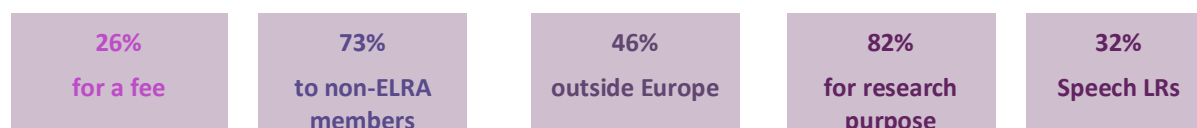
- ELRA-W0064: English, ELRA-W0066: French, ELRA-W0068: Greek
- ELRA-W0070: Italian, ELRA-W0072: Spanish
- ELRA-W0073 Quaero Old Press Extended Named Entity corpus
- ELRA-W0074 Amharic-English bilingual corpus

Evaluation Packages

- ELRA-E0041 CHIL 2007+ Evaluation Package

766

Language Resources distributed



Where are our clients in 2013?



↑ The tags show the countries.

Free Resources in 2013

Anticipating users' expectations, ELRA has decided to offer a large number of resources for free for Academic research use. Such an offer consists of several sets of speech, text and multimodal resources that are regularly released, for free, as soon as legal aspects are cleared. Whenever this is permitted by our licences, please feel free to use these resources for deriving new resources and depositing them within the ELRA catalogue for community re-use.

The first set of free LRs was released in May 2012, then early 2013, ELRA announced the release of the second set for the Academic research community (below). The list of the free LRs is available from www.elra.info.

[ELRA-E0003 TC-STAR 2005 Evaluation Package - ASR Spanish](#)

[ELRA-E0012-01 TC-STAR 2006 Evaluation Package - ASR Spanish - CORTES](#)

[ELRA-E0012-02 TC-STAR 2006 Evaluation Package - ASR Spanish - EPPS](#)

[ELRA-E0026-01 TC-STAR 2007 Evaluation Package - ASR Spanish - CORTES](#)

[ELRA-E0026-02 TC-STAR 2007 Evaluation Package - ASR Spanish - EPPS](#)

[ELRA-S0238 MIST Multi-lingual Interoperability in Speech Technology database](#)

[ELRA-S0268 UPC-TALP database of isolated meeting-room acoustic events](#)

[ELRA-W0031 GeFRPaC - German French Reciprocal Parallel Corpus](#)

[ELRA-W0017 MULTTEXT JOC Corpus](#)

[ELRA-W0018 ARCADE/ROMANSEVAL corpus](#)

[ELRA-S0083 ISLE Speech Corpus](#)

[ELRA-W0059 LT Corpus](#)

[ELRA-W0060 PTPARL Corpus](#)

[ELRA-W0061 CINTIL-DependencyBank](#)

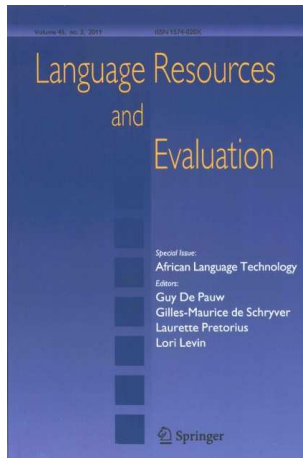
[ELRA-W0062 CINTIL-DeepBank](#)

[ELRA-W0013 TSNLP \(Test Suites for NLP Testing\)](#)

In line with this strategy and spirit, ELRA will continue to negotiate and enter into agreements to offer more free resources. Follow our announcements for this purpose.

To obtain the free LRs, you should
[download the End-User agreement from www.elra.info](http://www.elra.info)
and follow the instructions given in the agreement

Language Resources & Evaluation Journal



The LRE Journal, published by Springer, is the first publication devoted to the acquisition, creation, annotation, and use of LRs, together with methods for evaluation of resources, technologies, and applications. The ELRA members are granted complimentary access to the journal through the society on the condition that they subscribe to the publisher's table-of-contents alert service.

In 2013, the following regular issues, edited by Nicoletta Calzolari and Nancy Ide, have been published:

- Volume 47, n° 1, Spring 2013: Special Issues: **Collaboratively Constructed Language Resources** and **Analysis of short texts on the Web**
- Volume 47, n° 2, Summer 2013
- Volume 47, n°3, Fall 2013: Special Issues: **Computational Semantic Analysis of Language: SemEval-2010** and **Wordnets and Relations**
- Volume 47, n°4, Winter 2013

For LREC editions, the LREC organisers and the JLRE Editors have agreed with Springer on some special conditions for subscription to the JLRE to be offered to LREC participants.

Some Key Projects



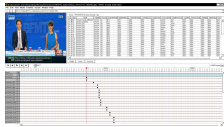
The **MLI Support Action** is working to deliver the strategic vision and operational specifications needed for building a comprehensive European MultiLingual data & services Infrastructure, along with a multi-annual plan for its development and deployment, and foster multi-stakeholders alliances ensuring its long term sustainability. More information can be found on the following website: <http://mli-project.eu>.



Focusing on Medical Information analysis and retrieval, the project KHRESMOI which stands for *Knowledge Helper for Medical and Other Information users* addresses both challenges of trustworthiness and complexity levels in online health information. The project is funded under the EU 7th Framework Programme ICT theme. The consortium, consisting of 12 partners from 9 European countries, will develop a web-based multi-lingual multi-modal search and access system for biomedical information and documents over the next four years. The system will allow querying in several languages, in combination with visual image queries. It will return translated document summaries linked to the original documents. This project continues under internal funds.



META-SHARE is one of the objectives within META-NET, it aims to be “a sustainable network of repositories of language data, tools and related web services documented with high-quality metadata, aggregated in central inventories allowing for uniform search and access to resources.” (<http://www.meta-net.eu/meta-share>). In its aim to achieve such a network, work is being carried out for the specification of a metadata schema which builds upon currently available schema, knowledge and expertise and provides a unified schema capable of handling the requirements of the community. These requirements comprise both the description of Language Resources and that of tools or technologies. To allow an easy understanding of the META-SHARE licences, ELRA is developing a legal Wizard to help the right-holders share/distribute their resources under the right licence. This will be also usable by users to better understand their rights and duties.



REPERE consists in an evaluation campaign for multimedia information processing systems. More specifically, it deals with person recognition within TV programs. On the organization side, two partners are funded by the DGA (Direction Générale de l'Armement): the LNE (Laboratoire national de métrologie et d'essais) organizes the campaign while ELDA is producing the evaluation data : acquisition of broadcasted programs, audio and text transcription, Named Entities annotation, head localization, head description and identification...



MAURDOR focuses on the automatic processing of digital documents. It attempts to improve processing technologies for handwritten and typewritten documents in French, English and Arabic, looking into various formats/quality (scanned, faxed, etc.) documents and having a special emphasis on efficient OCR solutions.



PEA-TRAD is a French national project which aims at developing Speech to Speech translation technology for several language directions and covering languages like Arabic, Chinese, English, French and Pashto. Moreover, PEA-TRAD has ambitiously targeted a variety of domains such as web-based text data, blogs, mail, and broadcast news.

ISLRN

International Standard Language Resource Number

<http://www.islrn.org>

After having reviewed existing identification schemas, ELRA and a large number of HLT key-players (NLP12) have concluded that there is a need to establish a specific identifier for Language Resources (LRs). Hence, we are pleased to introduce here, the **International Standard Language Resource Number (ISLRN)**. This new identification scheme is meant to provide LR with unique identifiers using a standardised nomenclature. It will ensure that LR are correctly identified, and consequently, recognised with proper references for their usage in applications within R&D projects, product evaluation and benchmarking, as well as in documents and scientific papers. Moreover, this is a major step in the networked and shared world that Human Language Technologies (HLT) has become: unique resources must be identified as such and meta-catalogues need a common identification format to manage data correctly. Therefore, LR should make use of identical identification schemes regardless of their representations, types and physical locations (on hard drivers, Internet or Intranet). ISLRN is supported by ELRA, LDC and AFNLP/Cocosda.

FLaReNet Follow-up



FLaReNet is a project funded by the European Commission under the eContentplus programme, a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable. The project is now finished but ELRA agreed to take over the management of the national contact points. In addition, ELRA agreed to work on the idea to host and maintain a repository of best practices and standards

LREC, the Language Resources and Evaluation Conference










The Language Resources and Evaluation Conference is an international scientific event which aims at providing an overview of the state of the art, exploring new R&D directions and emerging trends, exchanging information regarding Language Resources and their applications, carrying out the evaluation of methodologies and tools, on-going and planned activities, industrial uses and needs, requirements coming from the e-society, with respect to both policy issues and technological and organisational ones.

LREC provides a unique forum for researchers, industrial players and funding agencies from across a wide spectrum of areas to discuss problems and opportunities, find new synergies and promote initiatives for international cooperation in support of investigations in language sciences, progress in language technologies and development of corresponding products, services, applications, and standards.

In 16 years LREC which is organized every other year has become the major event on Language Resources and Evaluation for Human Language Technologies.

The conference usually covers a full week and the programme is organised around parallel oral and poster sessions during the main conference; 4 days before and after the conference will be dedicated to workshops and tutorials.

Since 1998, the LREC conference has been organized every other year with an increasing success. The following is a brief overview of all the places that have welcomed LREC:

		<p>LREC 2014 Reykjavik to come www.lrec-conf.org/lrec2014</p>			
<p>9th</p>					
<p>8th</p>	<p>LREC 2012 Istanbul 1200+ participants www.lrec-conf.org/lrec2012</p>		<p>4th</p>	<p>LREC 2004 Lisbon 950 participants www.lrec-conf.org/lrec2004</p>	
<p>7th</p>	<p>LREC 2010 Malta 1200+ participants www.lrec-conf.org/lrec2010</p>		<p>3rd</p>	<p>LREC 2002 Las Palmas 730 participants www.lrec-conf.org/lrec2002</p>	
<p>6th</p>	<p>LREC 2008 Marrakech 1100 participants www.lrec-conf.org/lrec2008</p>		<p>2nd</p>	<p>LREC 2000 Athens 600 participants www.lrec-conf.org/lrec2000</p>	
<p>5th</p>	<p>LREC 2006 Genoa 800 participants www.lrec-conf.org/lrec2006</p>		<p>1st</p>	<p>LREC 1998 Granada 510 attendees www.lrec-conf.org/lrec1998</p>	



MAIN CONFERENCE: 28-29-30 MAY 2014 **WORKSHOPS and TUTORIALS:** 26-27-31 MAY 2014

<http://www.lrec-conf.org/lrec2014> - Follow us on Twitter: @LREC2014

The format of the 9th edition of LREC is a bit different from the previous editions, with the post-conference workshops and tutorials shortened by one day (no Sunday). The selection of the workshops and tutorials has been done accordingly: 9 tutorials and 22 workshops have been accepted.

The format of the Main Conference remains unchanged (on 3 days) although the number of submissions is steadily increasing for each edition and has reached 1225 abstracts (vs 1027 in 2012 and 936 in 2010).

Finally, LREC 2014 will be held under the prestigious patronage of the UNESCO and with the support of Madame Vigdís Finnbogadóttir, former President of Iceland and UNESCO Goodwill Ambassador for Languages.

CONFERENCE TOPICS

- Issues in the design, construction and use of LRs: text, speech, multimodality
- Exploitation of LRs in systems and applications
- Issues in LT evaluation
- General issues regarding LRs & Evaluation

LREC 2014 HOT TOPICS

- Big Data, Linked Open Data, LRs and HLT
- LRs in the Collaborative Age

PROGRAMME

The Scientific Programme will include invited talks, oral presentations, poster and demo presentations, and panels, in addition to a keynote address by the winner of the Antonio Zampolli Prize.

Join ELRA

The annual membership fees are shown in the **Fees** column.

A Fidelity Programme has been implemented to reward the faithful members. The **Miles** column displays the number of miles (1 mile=1€) which is allocated to each member joining the association and/or remaining an ELRA member. Miles can be used to purchase resources, pay the membership fees or the registration to LREC.

	Fees (€)	Miles
Non-profit making organisations	750	200
European small/medium-sized companies (< 50 employees)	1000	250
European profit making organisations (>= 50 employees)	1500	375
Non-European profit making organisations	5000	1250

ELRA • 9, rue des Cordelières, 75013 Paris • France
Telephone: +33 1 43 13 33 33 • Fax: +33 1 43 13 33 30
info@elda.org • www.elra.info • www.elda.org