



# Annual Report 2015

ELRA • 9, rue des Cordelières, 75013 Paris • FranceTelephone:  
+33 1 43 13 33 33 • Fax: +33 1 43 13 33 30 [info@elda.org](mailto:info@elda.org) •  
[www.elra.info](http://www.elra.info) • [www.elda.org](http://www.elda.org)

## 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 involvement in the **CEF.AT** 2-year initiative launched in Spring 2015 which is part of the Connecting Europe Facility programme, and focus on public data for automated translation in the all European Member States plus Iceland and Norway
- The strong involvement in Language Resources **production projects**, as well as the participation to **HLT collaboration projects** such as MLI and CRACKER.
- The strong focus on development work:
  - New ELRA web portal
  - ELRA-SHARE, the new catalogue of LRs
  - LRE-Map, with the redesign of the interface and improved search tool.
  - The ISLRN portal
  - The Language Grid platform
- The celebration of the association's **20th anniversary** in October 2015, in Dubrovnik.
- The preparation of the 10<sup>th</sup> edition of **LREC** in Slovenia

"We continue to vision 2016-2020 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



## The ELRA Board

### Board Officers

**Nicoletta Calzolari**  
President



ILC CNR, Italy

**Khalid Choukri**  
Secretary General



ELDA, France

**Joseph Mariani**  
Honorary President



LIMSI & IMMI, France

**Nick Campbell**  
Vice-President



TCD, Ireland

**Isabel Trancoso**  
Vice-President



INESC, Portugal

**Maria Gavrilidou**  
Secretary



ILSP, Greece

**Andrejs Vasiljevs**  
Treasurer



Tilde, Latvia

### Board Members

**Thierry Declerck**



DFKI, Germany

**Marko Grobelnik**



JSI, Slovenia

**Henk van den Heuvel**



CLST, The Netherlands

**Frédérique Segond**



Objet Direct, France

## 2015 at a glance

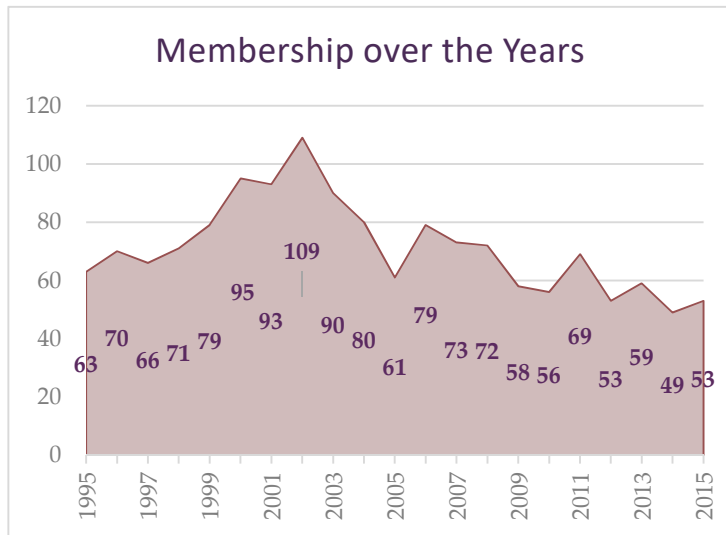
Jan	<ul style="list-style-type: none"><li>• Khalid Choukri chairs the <b>ISO/TC/SC35</b> Meeting in Copenhagen on January 18-19.</li><li>• ELRA attends the <b>LINDAT Clarin</b> Meeting in Prague on January 29.</li></ul>
Feb	<ul style="list-style-type: none"><li>• The <b>CRACKER</b> Kickoff Meeting takes place in Berlin on February 10.</li><li>• The Conference <b>TLRF 2015</b> (Technologies pour les Langues Régionales de France) takes place on February 19.</li></ul>
Apr	<ul style="list-style-type: none"><li>• The <b>CEF.AT</b> Kickoff Meeting takes place in Luxemburg on April 22.</li><li>• ELRA attends the Riga Summit on <b>Multilingual Digital Single Market</b> in Latvia on April 27-29.</li></ul>
May	<ul style="list-style-type: none"><li>• A ELDA-ILSP Joint Meeting on <b>Meta-Share Catalogue</b> takes place in Athens on May 11-13.</li><li>• ELRA attends the <b>MLi</b> General Meeting on May 27.</li></ul>
Jun	<ul style="list-style-type: none"><li>• A preliminary visit to Portoroz (Slovenia) on the <b>LREC 2016</b> site on June 3-6.</li></ul>
Aug	<ul style="list-style-type: none"><li>• Khalid Choukri chairs the <b>ISO/SC35 Plenary</b> Meeting in Hokkaido from August 31 to September 4.</li></ul>
Sep	<ul style="list-style-type: none"><li>• ELRA sponsors the 15<sup>th</sup> edition of <b>INTERSPEECH</b> in Dresden on September 6-10 and runs a booth during the conference.</li></ul>
Oct	<ul style="list-style-type: none"><li>• ELRA organizes a workshop to celebrate its <b>20<sup>th</sup> Anniversary</b> in Dubrovnik on October 12-14.</li><li>• ELRA attends the conference <b>ICT2015</b> in Lisbon on October 20-22.</li><li>• Khalid Choukri chairs the <b>ISO JTC 1st Plenary Meeting</b> in Beijing on October 26-31.</li></ul>
Nov	<ul style="list-style-type: none"><li>• The <b>Language Grid Workshop</b> takes place in Paris on November 12-13.</li><li>• ELRA attends the <b>EDF-MLi</b> meeting in Luxemburg on November 16-18.</li><li>• ELRA attends the <b>1<sup>st</sup> CEF LRB</b> Meeting in Berlin on November 19-20.</li><li>• ELRA attends the <b>Future and Emerging Trends in Language (FETLT2015)</b> in Sevilla on November 19-20.</li><li>• <b>ELRA Annual General Assembly</b> takes place in Paris on November 24.</li></ul>
Dec	<ul style="list-style-type: none"><li>• ELRA attends the <b>MT@EC</b> Meeting in Brussels on December 4.</li><li>• ELRA attends the <b>CEF.AT</b> Meeting on LRs in Frankfurt on December 16.</li><li>• ELRA attends the <b>CEF LINDAT</b> Meeting in Prague on December 17.</li></ul>

## ELRA Members

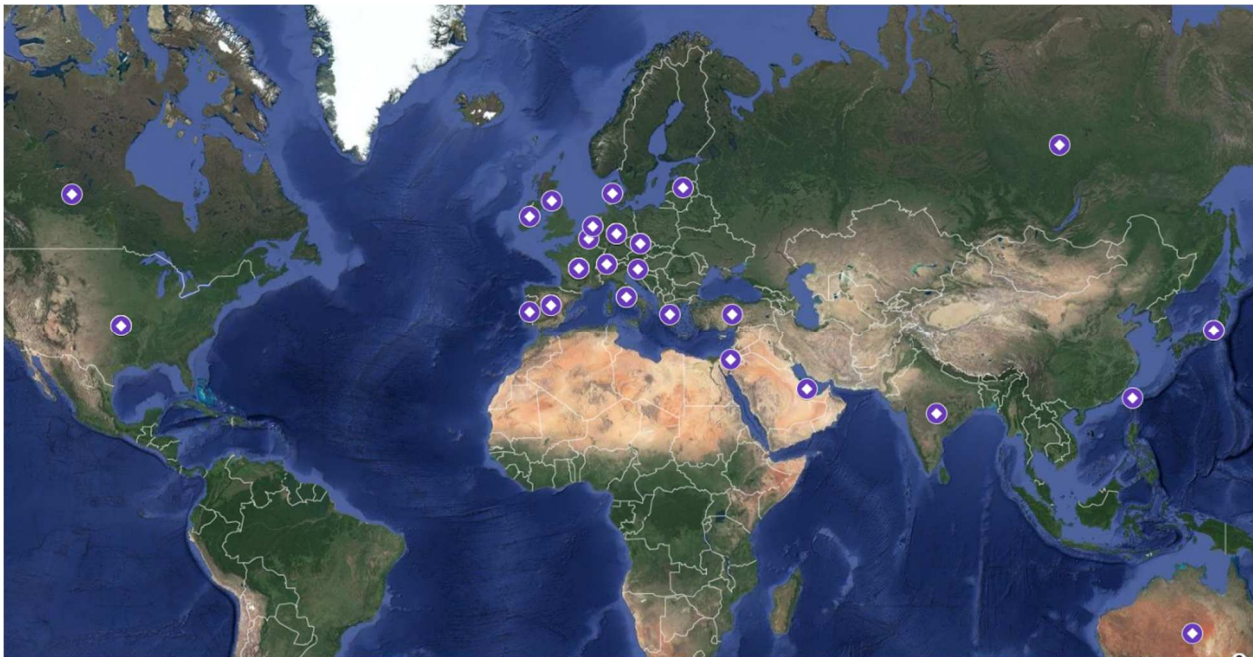
ELRA membership scheme is **institution-based**. Any organisation, public or private, European or non-European, can join. However, full membership, with voting rights, is available only to organisations legally established in Europe, as per the article 5 of the Statutes.

### 2015 ELRA Membership at a glance

<b>53</b>	<b>total number of members</b>
<b>19</b>	<b>new members joined in 2015</b>
<b>34</b>	<b>Members having renewed their membership</b>
<b>6</b>	<b>members joined in 1995 and have renewed membership continuously ever since</b>
<b>European members 74%</b>	



### Where are our members?



## Service 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](http://www.elra.info) web site.
- Free online access to **Language Resources and Evaluation (LRE) Journal**, published by Springer, is provided to all ELRA members.

## 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](mailto:elra-ipr-helpdesk@elra.info).

## *ELRA License Wizard*

To allow an easy understanding of the various licenses that exist for the use of Language Resources (ELRA's, META-SHARE's, Creative Commons', etc.), ELRA has developed a License Wizard to help the right-holders share/distribute their resources under the appropriate license. It also aims to be exploited by users to better understand the legal obligations that apply in various licensing situations.

The **License Wizard** works as a web configurator that helps the user:

- select a number of legal features,
- obtain the user license adapted to their selection,
- define which user licenses they would like to select in order to distribute their Language Resources,
- integrate the user license terms into a Distribution Agreement that could be proposed to ELRA or META-SHARE for further distribution through the ELRA Catalogue of Language Resources.



Reach the License Wizard portal and find the license for your Language Resources:  
<http://wizard.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 2015, these are:

- ERRARE 2015
- IJCAI workshop on Adaptive NLP 2015
- Interspeech 2015
- Riga Summit 2015

In addition, ELRA continues to support the Linguist List and has renewed its subscription to EAMT as an Institutional Member.

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

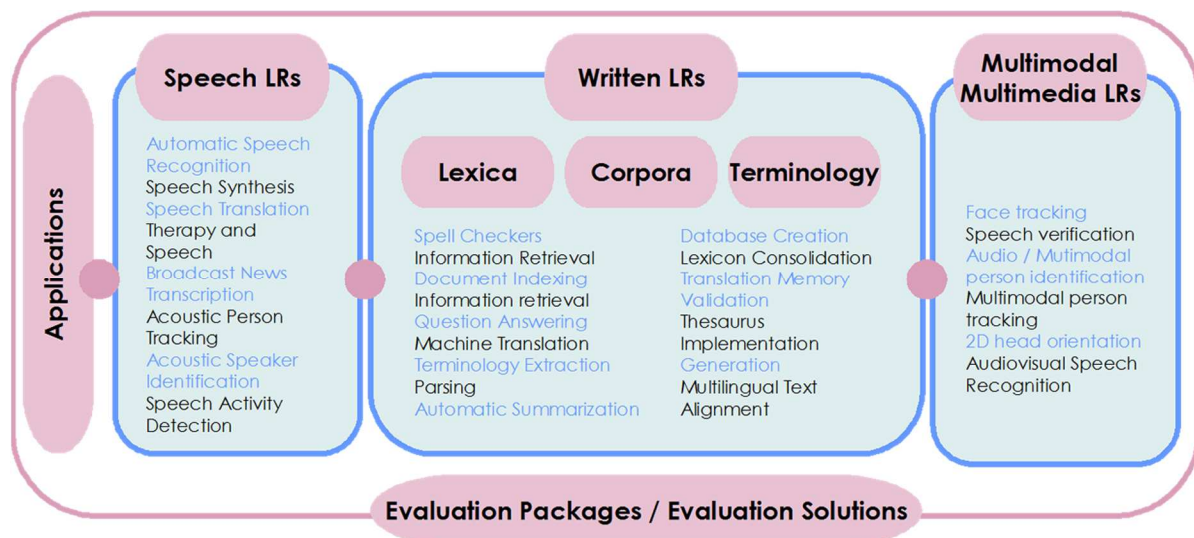
In addition to the promotion of language resources for the Human Language Technology (HLT) sector through market surveys, publication of a regular newsletter and organisation of LREC conference and other workshops, the activities of ELRA are organised around the following services:

- 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

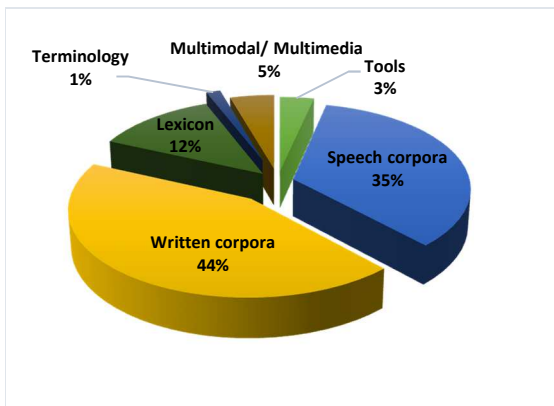
The Language Resources collected by ELRA are made available to the public through the ELRA Catalogue that can be accessed online at <http://catalog.elra.info>. The distribution of resources, which cover a wide variety of languages and belong to different modalities, is shown below:



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.



## 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.

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

## LRE Map

*The initiative for LR & Tools identification, the LRE Map*

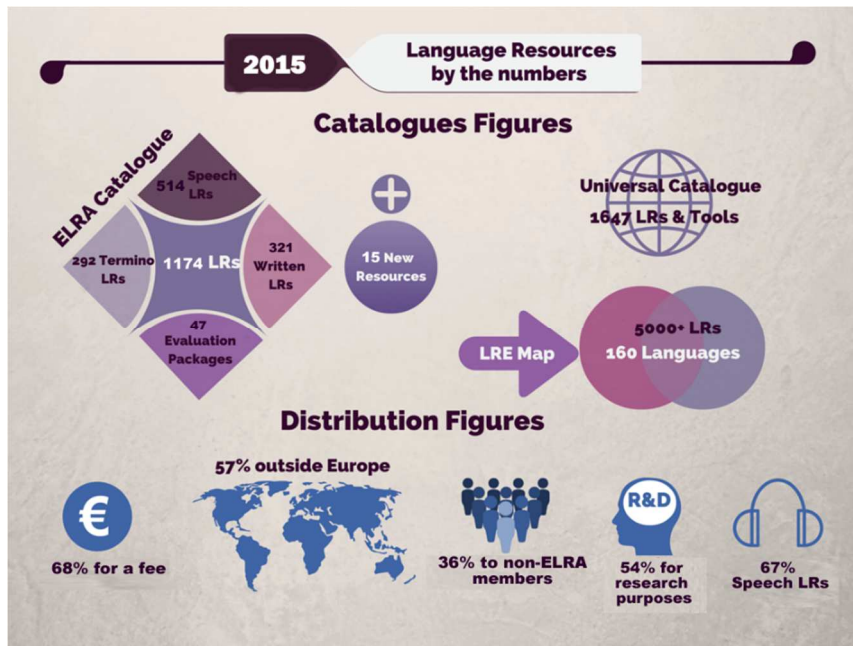
Initiated by ELRA and FlareNet at LREC 2010, the LRE Map is a mechanism intended to monitor the use and creation of language resources by collecting information on both existing and newly-created resources during the submission process. The feature has been so successful that it has been implemented also at other major conferences including COLING, IJCNLP, Interspeech, LTC, ACLHT, O-COCOSDA, RANLP, in addition to the *LRE Journal*.

The LRE Map feature is now part of the LREC standard submission processes for both the main conference and all the workshops. For LREC 2014, 1070 LR forms have been filled in (to compare with 900 LR forms in 2012). Globally, **5055 LR** resource type forms have been filled in within all the conferences, including LREC 2014, which have adopted the LRE Map.

At LREC 2014, as a new feature, the LRE Map was offering to the authors to share their resources by uploading them during the submission process.

All information on the LRE-Map can be found at <http://www.resourcebook.eu>.

# Language Resources: Catalogues & Distribution



## New LRs in 2015 & Distribution

### Written LRs

Khresmoi manually annotated reference corpus  
 ISLRN: 764-036-829-417-7, ELRA ID: ELRA-W0081

88milSMS. A corpus of authentic text messages in French  
 ISLRN: 024-713-187-947-8, ELRA ID: ELRA-W0082

deL1L2IM corpus  
 ISLRN: 339-799-085-669-8, ELRA ID: ELRA-W0083

Arboretum treebank  
 ISLRN: 025-729-182-451-2, ELRA ID: ELRA-W0084

ROCO Romanian journalistic corpus  
 ISLRN: 312-617-089-348-7, ELRA ID: ELRA-W0085

2006 CoNLL Shared Task - Ten Languages  
 ISLRN: 578-227-532-044-0, ELRA ID: ELRA-W0086

2006 CoNLL Shared Task – Arabic & Czech  
 ISLRN: 798-485-294-792-1, ELRA ID: ELRA-W0087

### Speech LRs

GVLEX tales corpus  
 ISLRN: 433-270-888-230-5, ELRA ID: ELRA-S0373

### Evaluation

CLEFeHealth 2014 Task 3 Evaluation Package  
 ISLRN: 725-020-897-275-7, ELRA ID: ELRA-E0043

REPERE Evaluation Package  
 ISLRN: 360-758-359-485-0, ELRA ID: ELRA-E0044

MAURDOR Evaluation Package  
 ISLRN: 364-018-517-901-2, ELRA ID: ELRA-E0045

### Monolingual Lexica

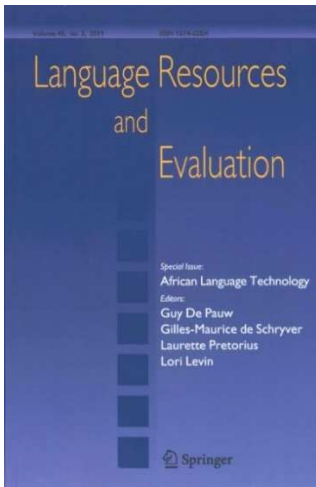
CEPLEXicon  
 ISLRN: 408-817-203-152-3, ELRA ID: ELRA-L0094

GLiCom Spanish Wordform list – Regular word-forms + verb-clitic combinations  
 ISLRN: 529-126-116-826-1, ELRA ID: ELRA-L0095-01

GLiCom Spanish Wordform list – Regular word-forms  
 ISLRN: 282-492-887-361-0, ELRA ID: ELRA-L0095-02

GLiCom Spanish Wordform list – Verb-clitic combinations  
 ISLRN: 064-262-883-254-3, ELRA ID: ELRA-L0095-03

# Language Resources and 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.

In 2015 the following regular issues, edited by Nicoletta Calzolari and Nancy Ide, have been published:

- Volume 49, n° 1. Regular issue
- Volume 49, n° 2. Regular issue
- Volume 49, n° 3. Regular Issue
- Volume 49, n° 4. Regular Issue

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



CRACKER is a Coordination and Support Action under the H2020 Programme from the European Commission. Launched beginning 2015, it is coordinated by DFKI.

CRACKER aims at providing planned coordination and support to the European Machine Translation research community, in the context of the the Digital Single Market setup. The principle behind this support action is that "sharing data, results and evaluation instruments" leads to a successful evolution and this collaborative work needs to go through the definition of priorities, planning, organisation, organisation of evaluations and then sharing of the outcome. For that purpose, CRACKER counts on a Consortium of experts covering all these needs and it builds upon established networks, infrastructures and evaluation campaigns. ELDA/ELRA participates in this project providing its expertise in all relevant aspects of the language resource sharing landscape.

More information can be found on the following website: <http://cracker-project.eu/>



The European Language Resource Coordination (ELRC) was launched to collect language resources for building machine translation systems for public service administrations across all EU Member States and Iceland and Norway which will better meet the everyday needs of public services across Europe. ELRC is an unprecedented public-sector data

collection effort aiming to provide CEF.AT with language and translation data (mono- and bi-lingual data) relevant to the daily needs of European national administrations. It aims not only to close the gap between the capabilities of the current MT systems offered by the EC to the national administrations and the actual, day-to-day requirements of national public services across Europe, but also to support Europe's national languages at the grass-roots level.

More information can be found on the following website: <http://lr-coordination.eu/>



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>.

### *Production Projects*

The following production projects were achieved in 2015 by the team at ELDA.

- Named-Entity corpora for evaluation (different domains) in Arabic, English and French
- Morphological and Named Entity Annotation in French, English and Arabic
- Tweet Data Annotation in French and German
- Reference Corpora for MT Evaluation from Arabic, Chinese, Persian and Russian to both English and French
- Spoken Language Production in American English

### International Standard Language Resource Number

The **International Standard Language Resource Number** (ISLRN), a unique and universal identification schema for Language Resources which provides Language Resources with unique identifier using a standardised nomenclature, ensures that Language Resources are correctly identified, and consequently, recognised with proper references for their usage in applications in R&D projects, products evaluation and benchmark as well as in documents and scientific papers.

The ISLRN Portal has been officially opened on April 3<sup>rd</sup>, 2014. Both ELRA and LDC have started to assign ISLRN to their resources. At the moment, the moderation is handled on ELRA's side. Since October 2015, LDC is moderating its own LRs in ISLRN. Since the official opening of the ISLRN Portal in April 2014 up to December 2015, 2061 numbers have been assigned, among which 1089 are provided by ELRA and 625 by LDC, the main providers.



Reach the ISLRN web site and request ISLRN for your own Language Resources:

<http://www.islrn.org>

## ***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 programme is organised around parallel oral and poster sessions during the main conference (3 days) and workshops and tutorials (2 days before and 1 day after the main conference). Since 1998, the LREC conference has been organized every other year with an increasing success.



**LREC 2016 Portorož**  
1200+ participants



**LREC 2014 Reykjavik**  
1200+ participants



**LREC 2012 Istanbul**  
1200+ participants



**LREC 2010 Malta**  
1200+ participants



**LREC 2008 Marrakech**  
1100 participants



**LREC 2006 Genoa**  
800 participants



**LREC 2004 Lisbon**  
950 participants



**LREC 2002 Las Palmas**  
730 participants



**LREC 2000 Athens**  
600 participants



**LREC 1998 Granada**  
510 participants

LREC 2016, 10<sup>th</sup> edition



**MAIN CONFERENCE:** 25-26-27 MAY 2016 - **WORKSHOPS and TUTORIALS:** 23-24-28 MAY 2016

<http://www.lrec-conf.org/lrec2016/lrec2016.htm>



@LREC2016

The format of the 10<sup>th</sup> edition of LREC is the one adopted at the Reykjavik edition: two days for the pre-conference workshops and tutorials and on day for the post-conference.

The format of the Main Conference remains unchanged (on 3 days)

#### PROGRAMME COMMITTEE

- **Nicoletta Calzolari**, *Conference chair*, CNR, ILC “Antonio Zampolli”, Pisa (Italy)
- **Khalid Choukri**, ELRA, Paris (France)
- **Thierry Declerck**, DFKI GmbH, Saarbrücken (Germany)
- **Marko Grobelnik**, Jožef Stefan Institute (JSI) (Slovenia)
- **Bente Maegaard**, CST, University of Copenhagen (Denmark)
- **Joseph Mariani**, LIMSI-CNRS & IMMI, Orsay (France)
- **Asuncion Moreno**, Universitat Politècnica de Catalunya, Barcelona (Spain)
- **Jan Odijk** – UIL-OTS, Utrecht (The Netherlands)
- **Stelios Piperidis**, Athena Research Center/ILSP, Athens (Greece)

#### 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

<http://www.elra.info/en/join-elra>



**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