Annual Report 2017

Word by the President & Secretary General of ELRA

Many strategic activities will be carried out in 2017 and 2018 by the ELRA Team (Board and staff). The highlights are as follows :

- **Preparation and organization of LREC 2018**

For the first time, the conference is held in Japan.

- **LRE-Map**

LRE-Map, a major feature of LREC, will offer a stand-alone version to help insert more resources by contributions. The main LRE-Map site will supply additional features to provide a sharper picture of LR availability and re-use.

- **Membership Policy**

The new plans and strategy for the evolution of the membership policy, to extend membership to all individuals attending LREC has been finalized and implemented. The ELRA statutes have been changed accordingly. Services around LRs to better respond to members' expectations have been redesigned.

- **LRL Committee, now Joint ELRA-ISCA SIGUL**

Work of the Less-resourced Languages Committee has continued. The workgroup on the Less-Resourced Languages has joined forces with ISCA's to become the Joint Special Interest Group ISCA & ELRA Special Interest Group Under-resourced Languages (SIGUL).

"We continue to vision 2017-2020 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. ELRA is also looking to more LR production activities on request."

Henk van den Heuvel

Khalid Choukri, Secretary General

ELRA Board

Board Officers

Henk van den Heuvel	Thierry Declerck	Maria Gavrilidou	Tatjana Gornostaja
President	Vice-President	Secretary	Treasurer
CLST, The Netherlands	DFKI, Germany	ILSP, Greece	Tilde, Latvia

Board Members

Gilles Adda	Nuria Bel	Antonio Branco	Marko Grobelnik	Simonetta Montemagni
LIMSI-CNRS, France	UPF, Spain	Univ Lisbon, Portugal	JSI, Slovenia	ILC-CNR, Italy

Honorary Presidents

Nicoletta Calzolari	Joseph Mariani
ILC-CNR, Italy	LIMSI-CNR, France

Secretary General

Khalid Choukri

ELDA, France

2017 at a Glance

Jan 17: ELRC+2 & ELRC+3 Inception Meetings, Luxemburg
Jan 31: Interministerial Group on Translation (GIT) Meeting, Paris (France)
Jan 31: ELRA Membership Drive Meeting, Paris (France)

January

February

Feb 6- 9:ISO SC35 Winter Meeting, Berlin (Germany)
Feb 15: First Call for Papers LREC 2018

Mar 28:CEF LRB 4th LRB Meeting, Berlin (Germany)

March

April

Apr 18 : ELRA Board Meeting, Paris (France)
Apr 20: ELRA-Vigdis Center MoU Signature, Reykjavik (Iceland)
Apr 25: InfoDay about Language Technologies, Madrid (Spain)

May 31: Final SMART 2014/1074 ELRC Meeting in Luxemburg
May 14-18 LREC 2018 Preliminary Trip to Miyazaki (Japan)

May

June

Jun 25-30 ISO/TC37 Plenary Meeting, Vienna (Austria)

Jul 4: ELRA Membership Drive Meeting, Paris (France)
Jul 19-20: ELRC+2 Workshop content Review Meeting with EC, Luxemburg

July

August

Aug 4 LREC 2018 Submission to Main Conference & Workshops opening
Aug 20-24 Interspeech 2017, Stockholm (Sweden)
Aug 23 SIGUL Special Session DigRev-URL, Stockholm (Sweden)

Sep 25: 1st Submission deadline for LREC 2018

September

October

Oct 2: 2nd Submission deadline for LREC 2018
Oct 2-7: ISO/IEC/JTC1 Plenary Meeting, Vladivostok (Russia)
Oct 13: ELRC+2 Workshop in Dublin (Ireland)
Oct 20 : ELRA Annual General Assembly

Nov 7:CEF ELRC 5th LRB Meeting, Brussels (Belgium)
Nov 7-8: CEF ELRC 3rd ELRC Conference, Brussels (Belgium)
Nov 21: CEF ELRI Kick-off Meeting, San Sebastian (Spain)

November

December

Dec 3-6: LREC 2018 PC Meeting, Paris (France)

Highlights on 2017 Activities

- **New Membership Policy**

A new membership policy has been worked out along the following lines: membership to be open to indvidual members and fidelity programme for institutional members replaced by a 3-year discount schema. Statutes have been changed accordingly and approved by the General Assembly. The implementation is starting on January 1st, 2018.

- **LREC 2018 in Miyazaki**

Part of 2017 has been dedicated to the organization of the 11th edition of LREC taking place in Miyazaki (Japan) in May 2018, in particular for all the activities related to the online submission of abstracts (with START), workshops and tutorial proposals. The Programme Committee meeting took place in Paris on 3-6 December 2017.

- **Language Resources**

New resources have been added to the catalogue on a regular basis. All of them are allocated an ISLRN to ensure that they are unique on one hand and improve both their persistency and the computation of their impact factor on the other hand.

- **Language Resources Projects**

In 2017, involvement in Language Resources production projects has remained strong. The CEF.AT initiative has been a strong focus throughout 2017 with the organization of the training workshops as well as the Language Resources collection. A follow-up initiative (ELRC+2 and ELRC+3) has been launched early 2017 and ELRI, a project aiming at setting up an infrastructure has begun late 2017.

- **Less-Resourced Languages Committee**

The workgroup on the Less-Resourced Languages has joined forces with ISCA's to become the Joint Special Interest Group ISCA & ELRA Special Interest Group Under-resourced Languages (SIGUL). Board elections took place at the end of 2017. A few events have been organized during the year, some as satellite of larger conferences like Interspeech.

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. As of January 2018, the membership scheme is changing to welcome individual members. All LREC 2018 participants are offered a 2-year subscription as part of their registration fees.

2017 ELRA Membership at a glance

**European Members
75%**

Members, incl. free membership

63 70 66 71 79 95 93 109 90 80 61 79 73 72 58 56 69 53 59 49 53 50 40

1995 1997 1999 2001 2003 2005 2007 2009 2011 2013 2015 2017

8 new members joined in 2017   **32** members renewed their membership in 2017

6 members joined in 1995 & have renewed membership continuously ever since

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** web site. Other reports are available to the members. They include: "Sustainability modeling" (produced with FlaReNet) and "LRs Production guidelines" (both technical and legal practices, provided in French language).

- Free online access to **Language Resources and Evaluation (LRE) Journal**, published by Springer, is provided to all ELRA institutional members.

- From January 1st 2018, the ELRA fidelity program is discontinued. The miles earned can be spent by May 2018. The faithful members can benefit from discounts on the membership fees (shown below) if they renew their membership for consecutive years.

		Y1 (starting from 2018)	Y2	Y3 and following years
	Basic Fees	100%	15% discount	30% discount
Non-profit making organisation	€ 750,00	€ 750,00	€ 637,50	€ 525,00
European small/medium-sized companies (< 50 employees)	€ 1 000,00	€ 1 000,00	€ 850,00	€ 700,00
European profit making organisations (>= 50 employees)	€ 1 500,00	€ 1 500,00	€ 1 275,00	€ 1 050,00
Non-European profit making organisations	€ 1 500,00	€ 1 500,00	€ 1 275,00	€ 1 050,00

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**.

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 2017, these are:

- EACL17
- CHISTERA Workshop
- International Workshop on Spoken Dialogue System (IWSDS)
- YFRSW 2017
- EUROPHRAS 2017
- EFNIL Workshop Mannheim
- European Big Data Forum 2017

In addition, ELRA has subscribed to Linguist List and joined the Big Data Value Association.

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.

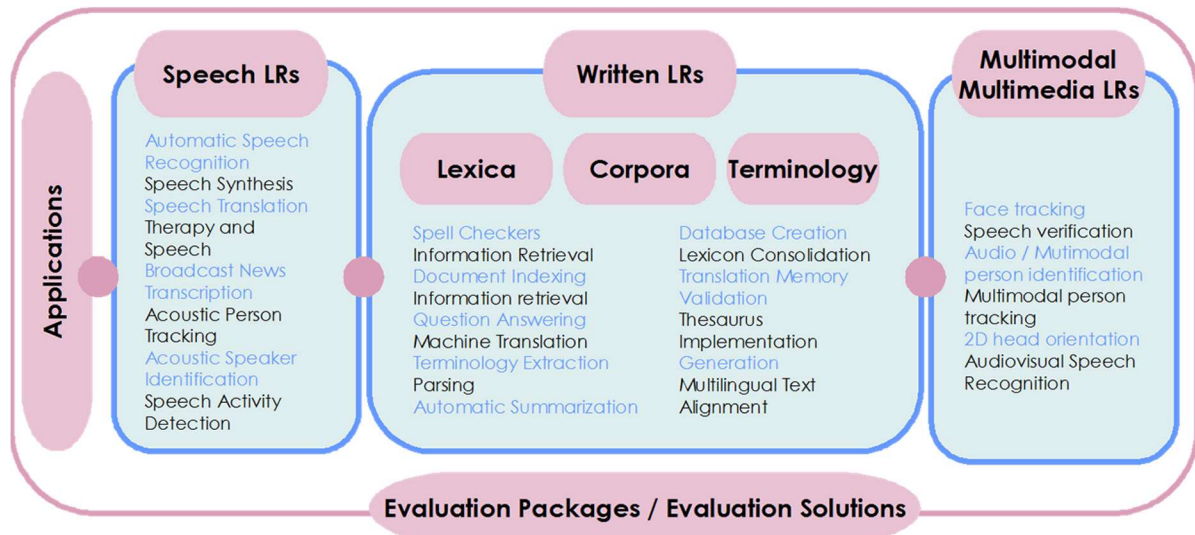## ELRA Activities and Projects

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.
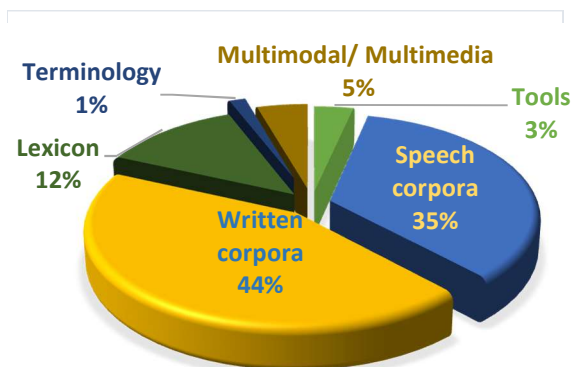
Language Resources Catalogue

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 collaborative: through a simple form, anyone 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

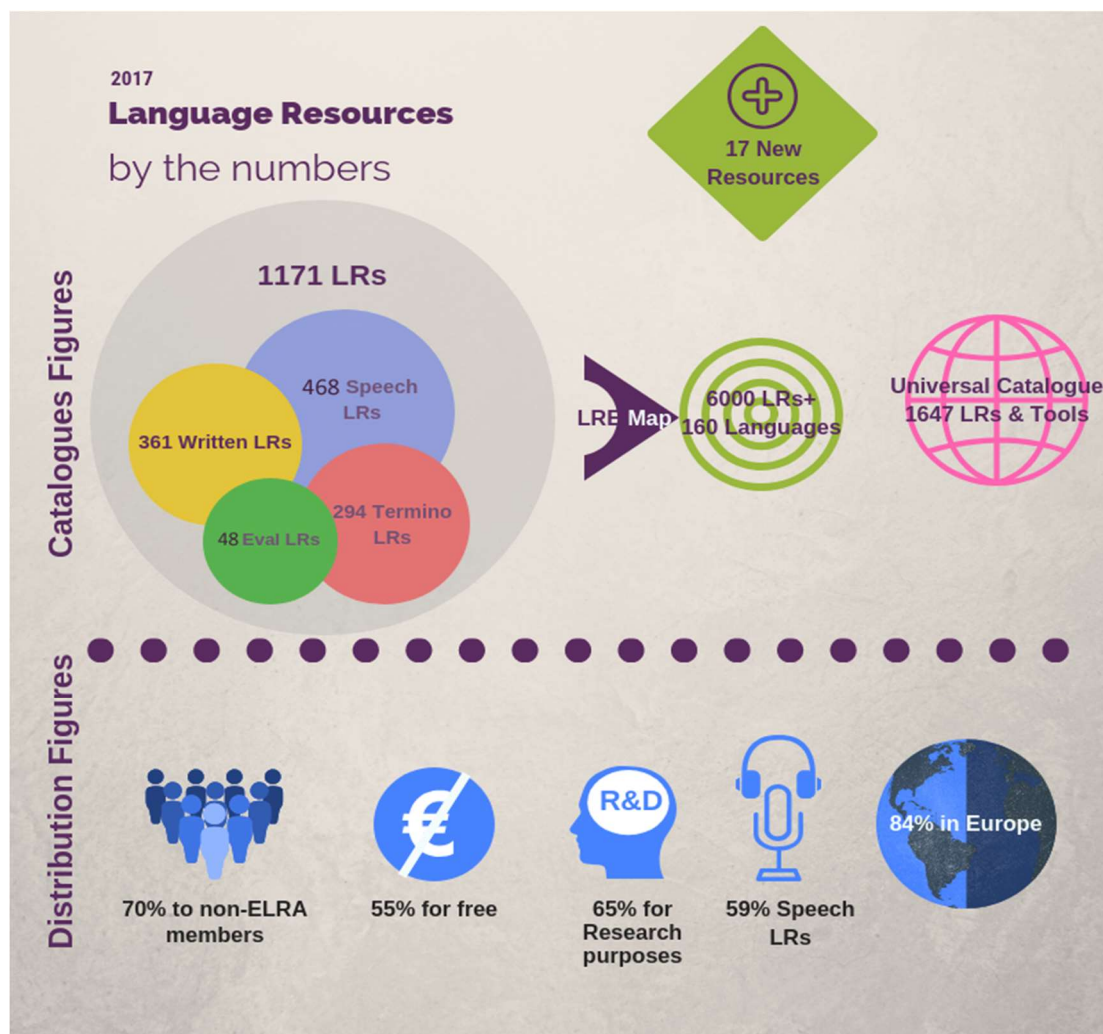## The initiative for LR & Tools identification, the LRE Map

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, **6143 LR** resource type forms have been filled in within all the conferences, including LREC 2016, which have adopted the LRE Map.

Since LREC 2014, the LRE Map is 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 2017 & Distribution

Speech Corpora	
Desktop/Microphone Speech	**Broadcast Speech Resources**
Accented English GlobalPhone, ISLRN: 574-579-221-841-3, ELRA ID: ELRA-S0389 Parallel EMG-Acoustic English GlobalPhone, ISLRN: 910-309-096-523-6, ELRA ID: ELRA-S0390 Persian Speech Corpus, ISLRN: 068-845-898-304-0, ELRA ID: ELRA-S0393	The FAME! Speech Corpus, ISLRN: 340-994-352-616-4, ELRA ID: ELRA-S0391
	Speech related Resources
	GlobalPhone Bulgarian Pronunciation Dictionary 260k entries (extended version), ISLRN: 799-402-906-876-5, ELRA ID: ELRA-S0388 Pashto phonetic lexicon, ISLRN: 186-827-325-462-6, ELRA ID: ELRA-S0392

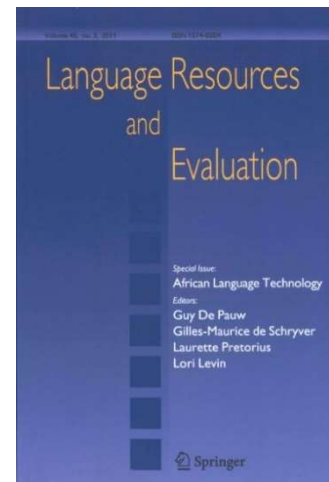| Written Corpora	
Written Corpora	**Monolingual Lexicons**
Danish Propbank (DPB), ISLRN: 213-212-351-142-5, ELRA ID: ELRA-W0117 English-Persian parallel corpus, ISLRN: 074-825-114-781-7, ELRA ID: ELRA-W0118 Helsinki Corpus of Swahili, ISLRN: 941-187-059-145-7, ELRA ID: ELRA-W0119 NUM 5M Mongolian written corpus, ISLRN: 492-817-146-504-9, ELRA ID: ELRA-W0120 2007 CoNLL Shared Task - Basque, Catalan, Czech & Turkish, ISLRN: 769-620-932-723-2, ELRA ID: ELRA-W0121 2007 CoNLL Shared Task - Greek, Hungarian & Italian, ISLRN: 270-733-242-642-3, ELRA ID: ELRA-W0122 2007 CoNLL Shared Task - Arabic & English, ISLRN: 505-782-255-628-8, ELRA ID: ELRA-W0123	Arabic dictionary of inflected words, ISLRN: 049-623-948-389-2, ELRA ID: ELRA-L0098 Arabic dictionary of inflected words with recognition of agglutinated clitics and inflection system, ISLRN: 963-860-792-289-9, ELRA ID: ELRA-L0099
	Evaluation Package
	ETAPE Evaluation Package, ISLRN: 425-777-374-455-4, ELRA ID: ELRA-E0046

Where are our customers?

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 2017 the following regular issues, edited by Nicoletta Calzolari and Nancy Ide, have been published:

- Volume 51, n° 1. Special Issue: Controlled Natural Language
- Volume 51, n° 2. Regular Issue
- Volume 51, n° 3. Regular Issue
- Volume 51, n° 4. Special Issue: Collaboration and Computing for Under-resourced Languages

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.

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 3rd, 2014. Both ELRA and LDC have started to assign ISLRN to their resources. The moderation is handled by ELRA and LDC. Since the official opening of the ISLRN Portal to December 2017, 2583 numbers have been assigned.

Reach the ISLRN web site and request ISLRN for your own Language Resources @ http://www.islrn.org

On-going Key Projects

The European Language Resource Coordination (ELRC) launched in 2015 to collect language resources for building machine translation systems for public service administrations across all EU Member States and Iceland and Norway ended up in May 2017.

Following the work of the European Language Resource Coordination (ELRC) action within CEF.AT, the European Commission has launched two further actions ELRC+2 and ELRC+3 under the same principles and also within the Connecting Europe Facility (CEF) Programme. The members of the European Language Resources Coordination Consortium (ELRC) remain DFKI, ELDA, ILSP and Tilde.

European Language Resource Coordination +2 (ELRC+2)

The goals of this 3-year project are to:
- set up and operate a repository to host Language Resources to support MT systems within CEF Automated Translation platform;
- set up and operate an intellectual property rights (IPR) support and clearance desk for Language Resources;

- complement and continue Language Resource coordination activities undertaken by the previous ELRC service contract, such as improving the availability of LRs held by the public sector, establishing an observatory for LRs across EU Member States and CEF associated countries and raising awareness among public data holders of the value of LRs for MT.

The project coordinated by DFKI, comprises ten tasks. ELDA is leading three of them:

- the technical helpdesk (T3),
- the legal helpdesk (T4)
- the IPR Clearance of 200 LRs (T5).

European Language Resource Coordination +3 (ELRC+3)

The main objective of ELRC+3 is to continue the ELRC's ongoing work in helping the EC obtain resources for the training and optimization of the CEF Automated Translation platform, for the CEF languages, and in domains of interest to the CEF Digital Service Infrastructures (DSIs). For that purpose, this action aims to identify, collect, clear, produce, process and make available further resources to the EC.

In this context, ELDA is leading the following activities:

- Adaptation of the existing ELRC database of sources, revising and customizing it for the new needs and requirements.
- Identification of licensing conditions and right holder(s) for the new resources.
- Dissemination activities, also in support of the ELRC+2 action.
- Anonymization of language resource databases: this will depend on the requirements of the language resource stakeholders and regulations on personal data protection.
- Validation of language resources and their metadata, which implies the quality evaluation of each deliverable language resource (both monolingual and parallel).
- Clearing of IPRs and other legal issues that may arise for the data collected.

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

ELRI, the European Language Infrastructure, is one of the latest actions in the framework of the CEF-TC-2016-3 programme.

The general objective behind ELRI is to deploy a network that allows for an accessible, secure and reliable sharing of LRs by different stakeholders at EU level. The focus is on language resources produced by translation centers and public institutions in Europe, aiming to render these data available for the European Commission (EC) so as to continue their work in the customization and improvement of their CEF Automated Translation platform to their service and to that of the Digital Service Infrastructures (DSIs) that will also share their data.

More information can be found on the following website: **http://www.elri-project.eu/**

Production Projects

The following production projects were achieved in 2017 by the team at ELDA.

- Sentiment Annotation in French Tweets
- Morphological and Named Entity Annotation in French, English and Arabic

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 20 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 2018 Miyazaki
1150 participants

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 2018, 11th edition

MAIN CONFERENCE: 9-10-11 May 2018 **WORKSHOPS & TUTORIALS:** 7-8-12 May 2018

http://www.lrec-conf.org/lrec2016/lrec2018.htm

@LREC2018

The format of the 11th 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** – CNR, Istituto di Linguistica Computazionale "Antonio Zampolli", Pisa - Italy (Conference chair)
- **Khalid Choukri** – ELRA, Paris - France
- **Christopher Cieri** – LDC, Philadelphia - USA
- **Thierry Declerck** – DFKI GmbH, Saarbrücken - Germany
- **Koiti Hasida** – The University of Tokyo, Tokyo - Japan
- **Hitoshi Isahara** – Toyohashi University of Technology, Toyohashi - Japan
- **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
- **Takenobu Tokunaga** – Tokyo Institute of Technology, Tokyo - Japan

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

A new membership policy adopted at the General Assembly in October 2017 has brought two main changes:

1. Individuals can now join the association as Individual Members
2. For Institutional Members, a new discount schema is replacing the miles-based membership fidelity programme which was terminated on December 2017.

Institutional Members Fees (for 1 year)	Fees (€)
Non-profit making organisations	750
European small/medium-sized companies (< 50 employees)	1000
European profit making organisations (>= 50 employees)	1500
Non-European profit making organisations	1500

Individual Members Fees (for 2 years)	Fees (€)
Individual Member (standard)	100
Individual Member (student))	50

The fees for 2018-2020 have been offered as part of LREC 2018 registration.

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