



ELRA ANNUAL REPORT 2022

Word by the President & Secretary General of ELRA

Many strategic activities have been carried out in 2022 by the ELRA Team (Board and ELDA staff). The highlights are as follows:

- **Work on LRs and enrichment of the ELRA Catalogue**

ELRA has continued to enrich the LRs catalogue. This process includes clearing all IPR issues before making resources available. Among the new LRs added to the catalogue in 2022: the Empathic Corpus of Interactions between Seniors and an Empathic Virtual Coach in Spanish, French and Norwegian and the 67 Speech Datatang resources distributed by ELRA.

- **New ISLRN website**

In 2022, ELRA has released a new version of ISLRN web portal. The website design has been reworked to improve the user experience: enhancement of navigation and search tools, extensive description of ISLRN, new icons.

- **LREC 2022 in Marseille**

The 13th edition of LREC took place in Marseille on June 20-25, 2022, as a hybrid edition, with 1800 registered participants (1/3 joined remotely) attending the main conference, the 34 workshops and 9 tutorials. The Antonio Zampolli Prize was awarded to Steven Bird.

- **Participation in major LR-related initiatives in Europe**

ELRA, through its agency ELDA, has continued to be actively involved in major LR and LT-related initiatives in Europe such as ELRC, ELG and ELE and bring its expertise in LRs management and relating legal issues.

- **Participation in LRs Production projects**

ELRA will continue to manage a number of production projects to support the development of LT/AI solutions, including building corpora for less-resourced language and hate speech detection in social media.

- **Support to SIGs**

Work of the Less-resourced Languages Committee has continued with SIGUL, including the organization of the SIGUL Meeting at LREC 2022. A new SIG on Spoken Language Translation has been setup in partnership with ACL and ISCA.

“We continue to vision 2020-2025 with emphasis on a renovation of ELRA policy and activities.

ELRA will continue to increase its visibility by forging new partnerships and strengthening current collaborations with CLARIN (for licences and distribution) or, UNESCO (LT4All and less-resources languages).

Renovation of ELRA such as the LRE Map will continue including the ELRA Catalogue, the Data Management Plan or the License Wizard.

Antonio Branco, President of ELRA

Khalid Choukri, Secretary General

MISSIONS & ORGANIZATION

ELRA AND ITS AGENCY, ELDA

Founded in 1995, ELRA is a non-profit organization whose main mission is to make Language Resources (LRs) for Human Language Technologies (HLT) available to the community at large.

To achieve this goal, ELRA carries out a wide variety of activities, including Identification & Distribution of LRs, Production & Validation of LRs, Technology Evaluation, Information Dissemination on HLT.

WHEN AND HOW DID IT START?

The history of ELRA starts in the late 1980s. Back then, the European Commission and some national agencies had encouraged and funded Language Engineering Research and Technology projects for nearly a decade, but the Language Resources developed for/within these projects were either lost or very difficult to re-use. The Research community and their sponsors therefore decided to act on this issue and ELRA was created in March 1995, with the mission of providing a **central clearing house for Language Resources**.

ELDA, THE OPERATIONAL BODY OF ELRA

ELDA, the Evaluations and Language resources Distribution Agency is ELRA's operational body. It is responsible for the **development and execution of ELRA's strategy**. Daily operations are supervised by its Managing Director, who reports to the ELRA Board. The activities cover the **practical and legal concerns** related to the distribution of resources, including **legal advice** provided to ELRA members and suppliers, the management of **user and distribution agreements** on behalf of ELRA. ELDA is also actively involved in **production, evaluation and dissemination** activities.

THE ELRA BOARD

ELRA's management is entrusted to a Board which consists of 9 elected members, including a member representing the individual members, who reflect the various dimensions of the Language Resources and Evaluation field. Each Board member is elected on a 2 year-term (and for 3 successive terms maximum) by an open vote of the ELRA members at the General Assembly.



<http://www.elra.info/en/about/elra/elra-board/>

THE ELRA MEMBERS

ELRA members include 40+ institutions (academic groups and research laboratories) and companies (industrial and commercial organisations) from all over the world (nearly all the countries of the European Union are represented) and 1000+ individual (researchers, users of LRs, etc.)

These institutions and companies belong to sectors like telecommunication, marketing, multilingual business, and deal with Human Language Technologies (HLT), especially speech, text and terminology.



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

MANAGEMENT OF ELRA

The ELRA Board will be substantially renewed in 2023 as 7 members have completed their mandate period. The nomination process took place late 2022, and the election by all ELRA members will happen early 2023.

ELRA BOARD

BOARD OFFICERS

Antonio Branco
President



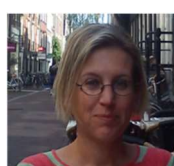
Univ Lisbon, Portugal

Simonetta Montemagni
Vice-President



ILC-CNR, Italy

Petya Osenova
Secretary



IICT-BAS, Bulgaria

Simon Krek
Treasurer



JSI, Slovenia

BOARD MEMBERS

Gilles Adda



LIMSI-CNRS, France

Nuria Bel



UPF, Spain

German Rigau



University of South Africa

Tatjana Gornostaja



Tilde, Latvia

Piek Vossen



VU Amsterdam, NL

HONORARY PRESIDENTS

Nicoletta Calzolari



ILC-CNR, Italy

Joseph Mariani



LIMSI-CNRS, France

SECRETARY GENERAL/ELDA CEO

Khalid Choukri



ELDA, France

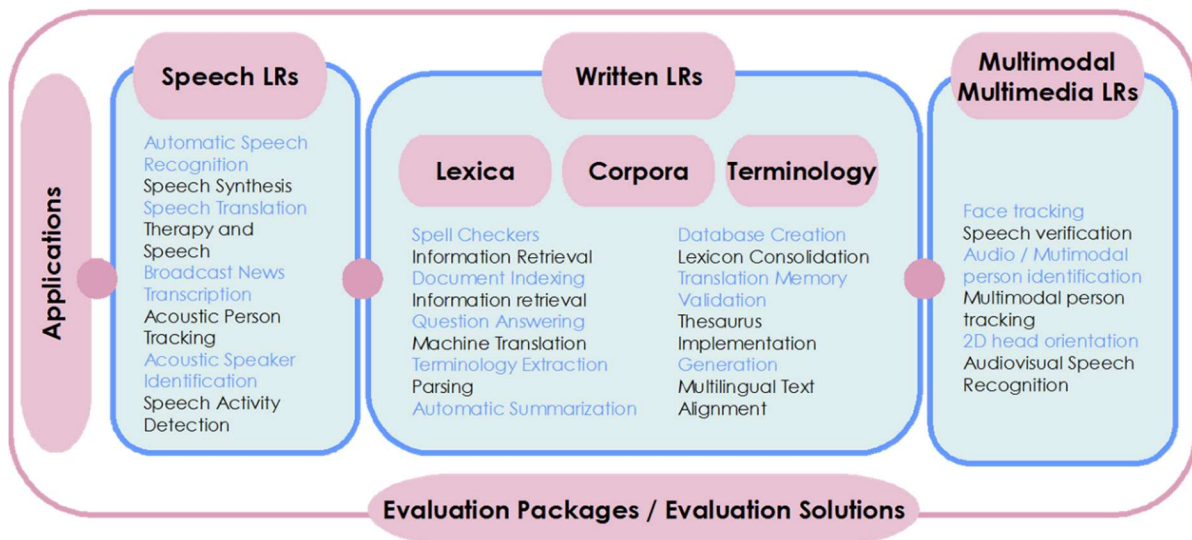
LANGUAGE RESOURCES

WHAT IS A LANGUAGE RESOURCE?

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.

The distribution of resources, which cover a wide variety of languages and belong to different modalities, is shown below:



DISTRIBUTION

LICENSING & DISTRIBUTION

As part of its distribution activities, ELRA has developed a **Licensing Model** in order to simplify the relationships between the producers/providers of Language Resources and the users. All agreements have been drafted with the support of lawyers.

To **distribute** a large range of resources and make them visible worldwide, ELRA has set up the ELRA Catalogue, a dynamic website, allowing the browsing and retrieval of detailed information on Language Resources.

Throughout the years, ELRA has developed an extended metadata set for the description of Language Resources. In the framework of its collaborative work, the ELRA Catalogue has been exported into OLAC and META-SHARE standards by means of associating the ELRA metadata for describing LRs with the OLAC and METASHARE metadata.



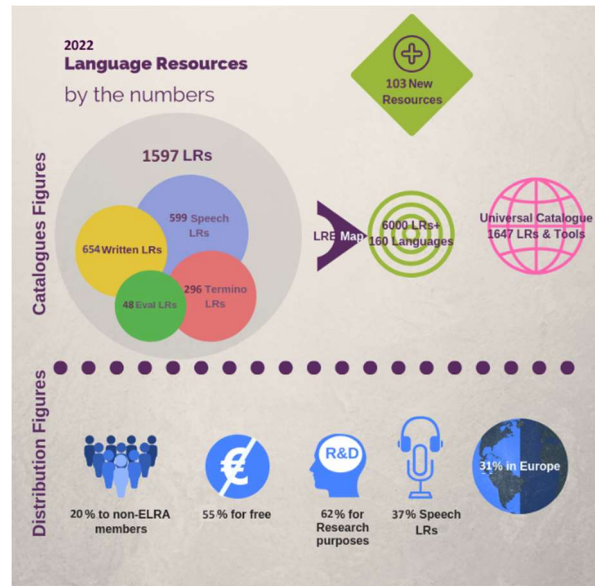
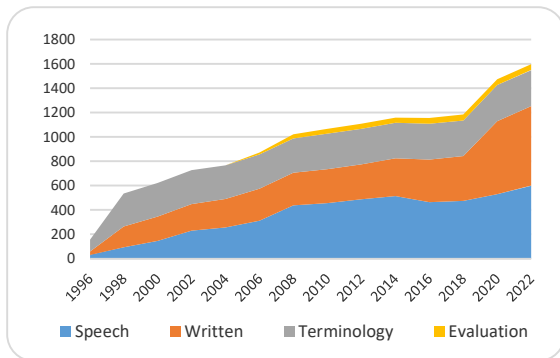
OLAC: <http://www.language-archives.org>

META-SHARE: <http://www.meta-share.eu>

CATALOGUES BY THE NUMBERS

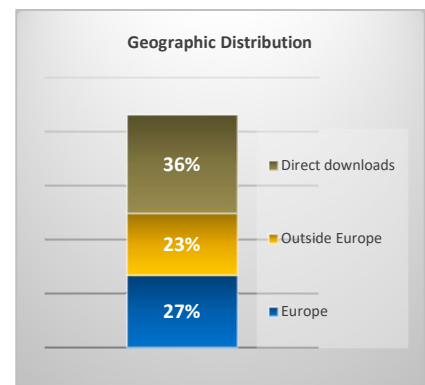
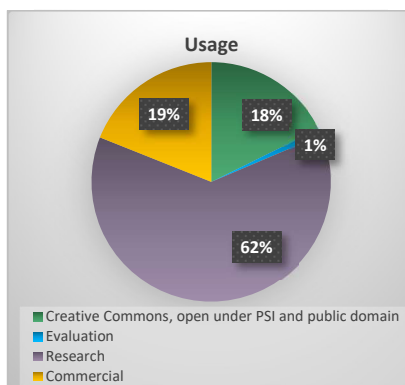
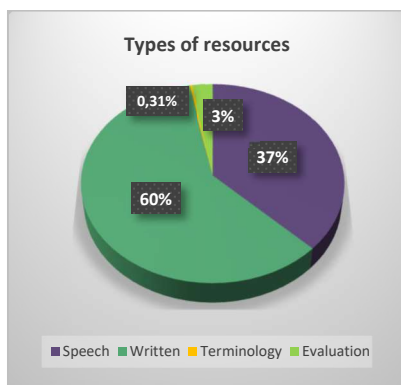
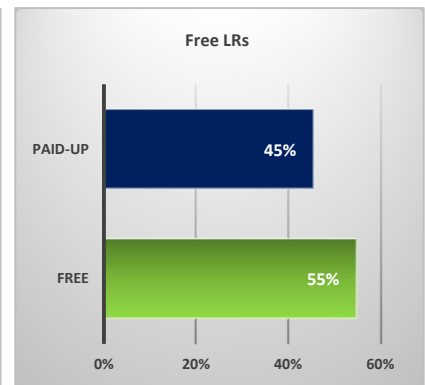
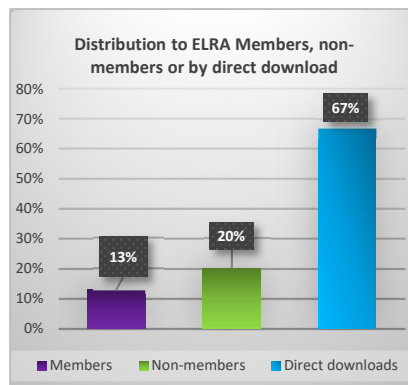
The Language Resources collected by ELRA are made available to the public through the ELRA Catalogue that can be accessed online at: <http://catalogue.elra.info>

The evolution of the LR types in the ELRA Catalogue is shown below.



As of end 2022, 1597 Language Resources are available in the ELRA catalogue of LRs.

OVERVIEW OF LRS DISTRIBUTED IN 2022



UNIVERSAL CATALOGUE
<http://www.universal.elra.info>

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.

LRE MAP
<http://www.resourcebook.eu>.

The initiative for LR & Tools identification

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. Globally, **9421** resource type forms have been filled since 2010.

PRODUCTION, A LONG-TERM EXPERTISE

Over the past 24 years, ELRA has compiled Language Resources (annotated corpora, corpora for MT evaluation, spoken LRs, etc.) in 30+ languages thanks to the large network of partners worldwide that ELRA has built overtime.

LRs can be produced in the framework of a contract with individual organizations or may be created within EU-funded projects, national projects or international cooperation projects, and may eventually become available in the ELRA catalogue

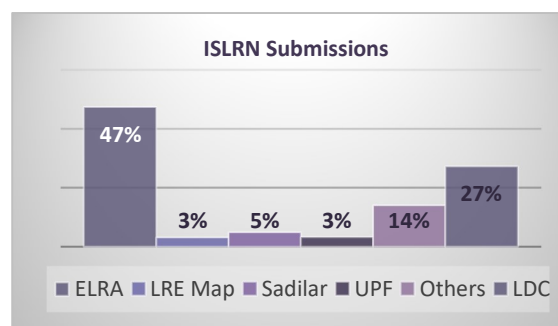
30+ technologies are covered including:

Automatic Speech Recognition	Monolingual textual Corpora	Spoken Language Translation	Multilingual Named Entities Corpora
Broadcast News Transcription	Aligned Corpora for Machine Translation	Spoken Language Understanding	Syntactic & Semantic annotations

ISLRN
<http://www.islrn.org>

The **International Standard Language Resource Number (ISLRN)** is a universal identification schema that provides a unique identifier using a standardized nomenclature. Each Language Resource (LR) with an ISLRN can therefore be recognized with its proper reference.

LREC 2022 endorses the need to uniquely identify LRs using ISLRN. All LRs cited in the accepted LREC papers will be assigned an ISLRN.



ELRA MEMBERS

MEMBERSHIP

ELRA membership scheme is two-fold: **institution-based and individual-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

2022 ELRA Institutional Membership at a glance



to welcome individual members. In 2022, **39 institutional members** and **886 individual members** have joined the association. For the individuals, most of them have subscribed through LREC registration and a few joined through the ELRA website.

Institutional Membership over the years



SERVICES TO ELRA 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.
- 30% discount on the membership fees by renewing their membership for 3 consecutive years or more.
- **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. Reports such Legal Issues on Web crawling or Market studies on LT in Europe are also shared with the members.
- Other reports are available to the members. Free online access to **Language Resources and Evaluation (LRE) Journal**, published by Springer, is provided to all ELRA institutional members every year.

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.



PROJECTS

PRODUCTION PROJECTS

ELDA is involved in several Language Resources production projects towards the production of annotated data. Production projects cover multiple languages and/or modalities. For the languages, ELDA can rely on a network of trusted partners worldwide to provide support for many languages, including less-spoken languages.

Whether commissioned by private customers or public institutions such as universities or research centers involved in research projects, the main objective of the corpora produced by ELDA is always to train general NLP systems, such as Machine

Translation engine, Speech processing system, or very specific systems such as Automatic subtitling system for deaf people. Some production projects are dedicated to supporting the detection of hate speech online.

Some production projects are confidential and do not appear in this document.

More details on the projects can be found @ <http://www.elra.info/en/projects/current-projects/production-lrs/>

Contact us @ info@elda.org if you need to produce LRs.

- Building a multilingual conversational telephone speech corpus
- Rosetta Project
- Tamasheq Project: transcription and translation of Tamasheq audio data for an Evaluation Campaign (Challenge IWSLT2022)
- POS-tagged Corpus in French and Spanish
- Audio Data Annotation for Speaker Identification
- Building a Speaker Identification Corpus for French
- Improve a conversational system
- Production of corpora for MT system
- Tweet Annotation for Hate speech detection

EUROPEAN INITIATIVES & PROJECTS

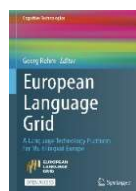
ELDA is involved in European initiatives and projects pertaining to the domain of Language Resources.

The **ELRC initiative** is dedicated to the collection of LRs in all the EU languages (plus Icelandic and Norwegian) to improve eTranslation, the EC Machine Translation Platform; primarily for public services in Europe, extended to SMEs.

The last meeting of the Language Resource Board concluding the initiative took place in Prague in November 2022.

All information on the European projects and how ELDA is contributing can be found @ <http://www.elra.info/en/projects/current-projects/>

The **ELG initiative** objective is to address the fragmentation in European LT business and research by establishing the European Language Grid as the primary platform for Language Technology in Europe (data and technology). The EU project ELG ran for 3.5 years and was completed successfully in June 2022.



In November 2022, the book *European Language Grid – A Language Technology Platform for Multilingual Europe* was published by Springer in the series « Cognitive Technologies ».

Other Projects

COVID-Mlia@Eval has been launched in June 2020 to improve Multilingual Information Access in the specific context of the COVID-19 pandemic. The areas addressed with COVID-Mlia are Data acquisition and Engineering, Information Extraction, Multilingual Semantic Search and Machine Translation.

More @ <http://eval.covid19-mlia.eu/>

DISSEMINATION

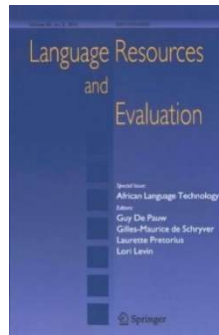
ELRA's main mission is to make Language Resources (LRs) for Human Language Technologies (HLT) available to the community at large. To achieve this

goal, ELRA carries out a wide variety of activities around LRs, including Information Dissemination on HLT, which are described in this section.

LANGUAGE RESOURCES AND EVALUATION JOURNAL

The LRE Journal, the official ELRA 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 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 2022, the following issues have been published. All the regular issues are co-edited by Nicoletta Calzolari and Nancy Ide:

- Volume 56, # 1, March 2022, Regular Issue
- Volume 56, # 2, June 2022, Regular Issue
- Volume 56, # 3, September 2022, Regular Issue
- Volume 56, # 4, December 2022, Regular Issue



<https://www.springer.com/journal/10579>

ELRA NEWSLETTER

In 2022, the ELRA Board has decided to produce a regular Newsletter to be sent regularly to all the ELRA members, institutional and individual.

The Newsletter's content is structured around the main topics :

- Language Resources
- Legal Issues
- ELRA/ELDA Projects
- Evaluation Campaigns
- Dissemination

In 2022, three issues have been published and disseminated using MailerLite.

- February 2022: [Issue #1](#)
- March 2022: [Issue #2](#)
- October 2022: [Issue #3](#)

All the issues are available from the ELRA site at <http://www.elra.info/en/dissemination/elra-newsletter/>

SPECIAL INTEREST GROUPS

<http://www.elra.info/en/sig/>

ELRA is actively involved in the following Special Interest Groups on the under-resourced languages on the one hand and on spoken language translation on the other hand.

ELRA is also contemplating setting up a SIG on Ethics and Privacy.

SIGUL

Created in April 2017, SIGUL is a joint Special Interest Group of the European Language Resources Association (ELRA) and of the International Speech Communication Association (ISCA).

SIGUL intends to bring together a number of professionals involved in the development of language resources and technologies for under-resourced languages in order to build a community supporting linguistic diversity through technology and ICT. but also commits to increase the lesser-resourced languages (regional, minority, or endangered) chances to survive the digital world through language and speech technology.



<http://www.elra.info/en/sig/sigul/>

SIGSLT

Setup in 2020 by ACL, ELRA and ISCA, the Special Interest Group on Spoken Language Translation (SIGSLT) shall promote interest in the research area of spoken language translation including simultaneous translation and interpretation, speech dubbing, speech-to-text translation, speech-to-speech translation, cross-lingual communication including paralinguistic, emotional or multimodal information and related areas. The SIG will also provide ACL, ELRA and ISCA members with a special interest in spoken language translation and its related areas with a means of exchanging news of recent research developments and other matters of interest in spoken language translation.



<https://bit.ly/3qWrtoz>

SIGUL ACTIVITIES IN 2022

Organization of the SIGUL 2022 Workshop @ LREC 2022

The **1st Annual Meeting of the ELRA/ISCA Special Interest Group on Under-Resourced Languages (SIGUL2022)** was held LREC 2022 in Marseille on June 24-25, 2022. 67 participants, including 51 attending in-person took part in the 1.5-day meeting.

SIGUL Board

- Chair and ISCA liaison representative: Sakriani Sakti (JAIST, Japan)
- Co-chair and ELRA liaison representative: Claudia Soria (CNR-ILC, Italy)
- Secretary: Maite Melero (BSC, Spain)

INTERNATIONAL DECADE OF INDIGENOUS LANGUAGES | 2022-2032



Back in 2019, ELRA took an active part in the organization of LT4All, the International Conference Language Technologies for All: Enabling Diversity and Multilingualism Worldwide at the UNESCO Headquarters in Paris. In Paris conference organized as part of the UNESCO.

In 2022, ELRA has joined the **Global Task Force** for the preparation of the International Decade of Indigenous Languages as a **permanent observer** and has taken part in several meetings, in addition to the **Official Launch event (livestream)** on 13 December 2022.

LREC, THE LANGUAGE RESOURCES & EVALUATION CONFERENCE

The LREC Conference has become a major event in the field of language engineering, and constitutes a milestone in the life of Human Language Technologies (HLT). It is organized by ELRA biennially since 1998 with the support of institutions and organizations involved in HLT, and brings together an increasing large number of participants working and interested in HLT.

A UNIQUE FORUM

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.



<http://www.elra.info/en/lrec>

LAST 10 EDITIONS OF LREC



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 2022, 20-25 MAY 2022

13th edition held at the Palais du Pharo, in Marseille (France)



<https://lrec2022.lrec-conf.org/en/> -  @LREC2022

- Hybrid edition (after cancellation of LREC 2020)
- 1810 participants (67% attended in person)
- 34 workshops: 1227 participants (79% in-person)
- 9 tutorials: by 229 participants (85% in-person)

Antonio Zampolli Prize awarded to **Steven Bird**, Northern Institute (Australia)

Proceedings available at: <http://www.lrec-conf.org/proceedings/lrec2022/index.html>

JOIN ELRA

Institutional Members Fees (for 1 year)	Year 1	Year 2	Year 3 (and following years)
	Basic Fee	15% discount	30% discount
Non-profit making organisations	750€	637,5€	525€
European small/medium-sized companies (< 50 employees)	1000€	850€	700€
European profit making organisations (>= 50 employees)	1500€	1275€	1050€
Non-European profit making organisations	1500€	1275€	1050€

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



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