

ELRA, the European Language Resources Association, is a worldwide leading player providing Distribution and Production Services to the Human Language communities since 1995.

The ELRA Catalogue of Language Resources, a large offer of Data for HLT

Over the last decades, ELRA has compiled a large list of resources that constitute its **Catalogue of Language Resources**. ELRA has negotiated distribution rights with the LR owners and made such resources available under fair conditions and a clear legal framework. Throughout the years, ELRA has developed detailed LR description forms and an extended metadata set for the description of Language Resources.

As of June 2021, the ELRA Catalogue of Language Resources counts **1381 Language Resources** shared over the following types of data:

- Speech and Speech-related LRs (eg. pronunciation dictionaries),
- Multimodal LRs,
- Written corpora,
- Monolingual and multilingual lexica,
- Terminological resources.



<http://catalog.elra.info>

Identification & Distribution of Language Resources

Free LRs

Anticipating users' expectations, ELRA has decided to offer a large number of resources for free for Academic research use. Such an offer will consist of several sets of speech, text and multimodal resources that will be regularly released, for free, as soon as legal aspects are cleared. The LRs released so far are mostly Speech LRs and Evaluation Packages for Speech technologies, as described below.

The Universal Catalogue

In addition to the resources available off the shelf from the ELRA Catalogue, ELRA has also worked on LR discovery and identification with a dedicated team that investigated and listed 1000+ existing and valuable resources and stored the information in the Universal Catalogue a list of resources that could be negotiated on a case by case basis.

We also provide **On-demand Identification Services**.



<http://universal.elra.info>

LRE Map, finding new ways through LRs

The LRE Map is an inventory of Language Resources and tools that were described in LREC papers. Such a huge inventory listed by the authors themselves constitutes a first “community-built” catalogue of existing or emerging resources, constantly enriched and updated at major conferences. It has been first introduced at LREC 2010 and used ever since at LREC and in other conferences.



<http://lremap.elra.info>

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  @ELRANews

ISLRN, an International Standard Language Resource Number

The **International Standard Language Resource Number (ISLRN)**, established by the NLP12 in 2013, 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 whether it is used in applications in R&D projects or in product evaluation and benchmark, or cited in documents and scientific publications.

The ISLRN does not replace local and specific identifiers, it is not a legal deposit nor an obligation, rather a best practice: a Language Resource distributed by several data centers with a “local number” or published in several meta-catalogues can still be allocated a unique ISLRN.



<http://www.islrn.org>

SIGUL, the Special Interest Group on Under- resourced Languages

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

Through its participation in the **Special Interest Group on Under-resourced Languages**, ELRA reasserts its active involvement in contributing to enhance the support for the languages with little or no technological support.

Since Language Resources and technologies represent a key component for any language-based technology, this special interest group intends to focus on the particular needs and requirements of less-resourced languages.



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

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  @ELRANews

Language Resources Production

The Production Services, a long-term expertise @ ELRA

Over the years, ELRA has compiled Language Resources in more than 25 languages. This could be achieved thanks to the large network of partners worldwide that ELRA has built overtime. ELRA remains involved in every stage of the Language Resources production and our end-to-end production services for Speech Language Resources include:

- Setup of the production environment, including the recording platforms
- Recruitment of speakers all over the world
- Management of the data collection team
- Collections of recordings in various setups: in cars, public places, offices, studio, etc.
- Transcriptions and annotations of audio data
- Validation

We have produced Language Resources for specific technologies which include (but are not limited to):

| | |
|-------------------------------|---------------------------|
| Automatic Speech Recognition* | Speaker Recognition |
| Broadcast News Transcription | Text To Speech synthesis |
| Spoken Language Translation | Acoustic Person Tracking |
| Spoken Language Understanding | Speech Activity Detection |

* (from prepared to spontaneous speech and conversational speech)