

ELRA's Consolidated Services for the HLT Community

Victoria Arranz, Khalid Choukri, Valérie Mapelli, H el ene Mazo

ELDA/ELRA

9 rue des Cordeli eres, 75013 Paris, France

E-mail: {arranz; choukri; mapelli;mazo}@elda.org

Abstract

This paper emphasises on ELRA's contribution to the HLT field thanks to the consolidation of its services since LREC 2012. Among the most recent contributions is the establishment of the International Standard Language Resource Number (ISLRN), with the creation and exploitation of an associated web portal to enable the procurement of unique identifiers for Language Resources. Interoperability, consolidation and synchronization remain also a strong focus in ELRA's cataloguing work, in particular with ELRA's involvement in the META-SHARE project, whose platform is to become ELRA's next instrument of sharing LRs. Since last LREC, ELRA has continued its action to offer free LRs to the research community. Cooperation is another watchword within ELRA's activities on multiple aspects: 1) at the legal level, ELRA is supporting the EC in identifying the gaps to be fulfilled to reach harmonized copyright regulations for the HLT community in Europe; 2) at the production level, ELRA is participating in several international projects, in the field of LR production and evaluation of technologies; 3) at the communication level, ELRA has organised the NLP12 meeting with the aim of boosting co-operation and strengthening the bridges between various communities.

Keywords: ISLRN, sharing LRs, interoperability, HLT support and cooperation, legal issues

1. Introduction

One big novelty in ELRA's latest developments stands in the establishment of the International Standard Language Resource Number (ISLRN), with the creation and exploitation of an associated web portal to enable the procurement of unique identifiers for Language Resources.

Interoperability, consolidation and synchronization remain also a strong focus in ELRA's cataloguing work. This has been concretised through the maintenance of various catalogues and repositories, as well as their mutual harmonization. Lately ELRA's involvement in the META-SHARE project, together with the exploitation of the resulting platform is pathing the way to new extensions of the platform so as to become ELRA's next instrument of sharing LRs. Sharing LRs also means opening access to such LRs. ELRA has reviewed its distribution policy and has increased its offer of free LRs since last LREC.

Cooperation is another watchword within ELRA's activities on multiple aspects. Cooperation at the legal level is one of those aspects, as ELRA has recently worked with the EC at identifying the gaps to be fulfilled in order to reach a more harmonized copyright law for the HLT community in Europe. Moreover, ELRA has carried on its participation in several international projects, in the field of Language Resource production and evaluation of technologies, thus keeping on contributing to the expansion of the field. Thanks to ELRA's experience in promoting and communicating in the field of Human Language Technologies, new cooperation action points could be highlighted. This could be done in particular thanks to the organisation of the NLP12 meeting that took place with the aim of boosting co-operation and strengthening the bridges between various communities.

2. Establishing the International Standard Language Resource Number (ISLRN)

2.1 Setting up of the ISLRN Infrastructure

12 major NLP organisations which gathered on November 18, 2013, at the first NLP12¹ Coordination Meeting, announced the establishment of the ISLRN, a Persistent Unique Identifier, to be assigned to each Language Resource. Experiment replicability, an essential feature of scientific work, would be enhanced by such unique identifier. Set up by ELRA, LDC² and AFNLP/Oriental-COCOSDA³, the ISLRN Portal aims to provide unique identifiers using a standardised nomenclature, as a service free of charge for all Language Resource providers.

The opening of the ISLRN Portal⁴ was officially announced on 3rd April 2014. To start with, Language Resources from both LDC and ELRA's catalogues were the first resources to be assigned a 13-digit ISLRN number.

For the time being, the overall activity is managed by ELRA, with the support of LDC. Later on, it will be supervised by a steering committee composed of representatives from participating organisations and it will be enlarged whenever necessary.

2.2 ISLRN Modus Operandi

The portal's modus operandi can be summarized as follows. Two types of external players may interact with the ISLRN Portal: Visitors and Providers.

1) Visitors of the portal may browse the web site and search for the ISLRN of a given Language Resource by its name or by its number if it exists. The functionalities that can be accessed by Visitors are:

¹ <http://www.elra.info/NLP12-Meeting.html>

² <https://www ldc.upenn.edu>

³ <http://www.afnlp.org> & <http://www.cocosda.org>

⁴ <http://www.islrn.org>

- Identify a Language Resource according to its ISLRN.
- Identify an ISLRN by the name of a Language Resource.

- Get information about ISLRN.
- View last 5 accepted resources.
- Sign up to become a provider.

2) Providers are registered and own credentials. They can request a new ISLRN for a given Language Resource. A provider has the possibility to become certified, after moderation, in order to be able to import metadata in XML format. The functionalities that can be accessed by Providers, once they have signed up, are:

- Log in.
- Request an ISLRN according to the metadata of a given resource.
- Request to become a certified provider so as to import XML files containing metadata.
- Import one or more metadata descriptions in XML to request ISLRN(s) (only for certified providers).
- Edit pending requests.
- Access previous requests.
- Contact a Moderator or an Administrator.
- Edit Providers' own profile.

An ISLRN request is handled by moderators within 5 working days.

3. Consolidating the ELRA Catalogues and LR Repositories

3.1 Harmonizing the ELRA Catalogue and other Repositories

Since 1995, the ELRA Catalogue⁵ has been growing constantly, starting with a simple list of identified data and evolving to be a major archive of Language Resources available off-the-shelf. Continuous effort has been provided in order to convince LR producers and owners to release their assets, so as to capitalize on them and to enhance the Human Language Science and Technology field, thanks to the exploitation of adapted material. To support this work, several actions have been undertaken to make the LRs available and visible widely.

Through its ongoing discussions with LR owners, ELRA has constituted a catalogue which now consists of nearly 1,100 LRs available for various types of HLT applications in more than 60 different languages. Recently, the number of LRs in the catalogue has slightly stagnated, not because of the lack of LRs but due to the fact that ELRA has done a bigger effort towards the consolidation of the current offer by paying more attention to the community expectations.

In parallel, ELRA has worked on making its catalogue interoperable with other repositories, to enable a wider access to the information on LRs. This work has been based on the users' expectations to see standardised

descriptions that better qualify the resources in terms of content and quality. Among the initiatives in this field, we can mention the OLAC (Open Language Archives Community)⁶ initiative. Since September 2009, the ELRA Catalogue is automatically harvested by the OLAC server. It was initially exported into OLAC standards by means of associating the ELRA metadata for describing LRs with the OLAC metadata. Since then, it can be viewed as an OLAC repository and achieves the maximum rate of 5 stars on the OLAC scale.

Considering the latest trends on easing the sharing of LRs, from both legal and commercial points of view, ELRA has taken a major role in META-SHARE⁷, a large European open infrastructure for sharing LRs, which was initiated in 2009 within META-NET⁸. It was officially launched in Berlin on January 25, 2013, and came standard with an updated repository management tool and standardization of metadata schema. Lately, ELRA has worked at consolidating the synchronized information between its catalogue and the META-SHARE platform. ELRA is also one of the managing nodes, whose task is to share metadata, create, host and maintain a central inventory which includes metadata-based descriptions of all Language Resources available in the distributed network. Each META-SHARE managing node effectively hosts a copy of the central inventory.

As a complementary service, ELRA created an antechamber of the ELRA Catalogue: the Universal Catalogue⁹. This catalogue offers a compilation of about 1,600 LRs worldwide which have been identified by the ELRA/ELDA team.

Moreover, ELRA has continued its identification work through the *LRE Map*¹⁰ initiative, launched at LREC 2010. This is implemented by collecting information on both existing and newly-created resources during the papers/abstracts submission process (Calzolari et al. 2010). At LREC 2012, nearly 4,000 LRs were identified and the work is ongoing at LREC 2014. Mining all the information provided by LREC 2012 authors has contributed to the revision of the list of values defined for several fields for this 2014 edition.

3.2 Focusing on free LRs

ELRA plays a leading role in the distribution and sharing of Language Resources within the Human Language Science and Technology Community. Over the last decade, ELRA has compiled a large list of resources that constitute its Catalogue of LRs. ELRA has negotiated distribution rights with the LR owners and made such resources available under fair conditions and a clear legal framework.

⁶ <http://www.language-archives.org>

⁷ <http://www.meta-share.eu>

⁸ <http://www.meta-net.eu>

⁹ <http://universal.elra.info>

¹⁰ <http://www.resourcebook.eu>

⁵ <http://catalog.elra.info>

Listening to the community has led ELRA to revise its distribution policy thoroughly. One of the main issues pointed out by the R&D community was the pricing policy. In order to adapt its offer to the widest community, ELRA has conducted several actions so as to adjust its prices to the research activities in need of resources. Further efforts have been dedicated to re-negotiate the pricing of LRs already distributed in the ELRA catalogue and thus get providers to agree on better pricing conditions.

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. Up to now, two sets were released, amounting 13 Evaluation Packages, 6 Speech Resources, 12 Written Lexica and Corpora¹¹.

In line with this strategy and spirit, ELRA continues to negotiate and enter into agreements to offer more free resources.

4. Evaluation and Evaluation-Oriented LR Production

ELRA continues to play an active role in the field of HLT evaluation, offering through ELDA a number of services that concern a) its implication in evaluation activities and campaigns, b) the production of LRs for evaluation, and c) the later distribution of the generated evaluation packages. These evaluation services cover a wide range of technologies, going from text to speech or multimodal technologies. In this regard, we should mention ELDA's participation in a number of projects and initiatives these past two years, some of them still ongoing, some others recently closed. These follow below, organised by the major technological areas.

4.1 Information Access

ELDA has collaborated in the evaluation of information access systems for years, very particularly in the *CLEF Initiative*¹² (Conference and Labs of the Evaluation Forum), which promotes the systematic evaluation of information access systems through experimentation on shared tasks, with an emphasis on multilingual and multimodal information. This initiative consists in evaluation labs conducting the evaluation activities and a peer-reviewed conference discussing the results from such activities and exploring all aspects around evaluation. The CLEFeHealth Lab organized a User-centred health information retrieval Shared task in 2013 which has resulted in an evaluation package that ELRA has made available¹³. This package contains all datasets used for the task, the queries used for these evaluations, the corresponding relevance assessments, the official

results obtained as well as the guidelines and working notes to help understand the whole procedure.

Following on this work, ELDA has also participated in the EU FP7-funded *PROMISE* Network of Excellence¹⁴, which has also been dedicated to the evaluation and benchmarking of complex multimedia and multilingual information systems, taking care of aspects such as access, curation, preservation, re-use and analysis, of the collected data.

*KHRESMOI*¹⁵ (another EU FP7-programme project) has been aimed at developing a multilingual multimodal search and access system for biomedical information and documents. The research flows into several open source components, which are integrated into an innovative open architecture for robust and scalable biomedical information search. ELDA participates in a number of activities within the project concerning LRs identification, IPR clearing, dissemination and exploitation, as well as evaluation. For instance, KHRESMOI has been the provider for the collection of medical-related documents available within the earlier-mentioned CLEFeHealth 2013 Evaluation Package (see CLEF's description above).

4.2 Machine Translation Technologies

The evaluation of Machine Translation (MT) and Speech-to-Speech Machine Translation (SST) systems has been one of ELRA/ELDA's strongest focus since its early evaluation days. These past few years, ELDA has looked into a variety of aspects around these technologies, being also involved in other related aspects.

For instance, ELDA has participated in *PANACEA*¹⁶ (which stands for "Platform for Automatic, Normalized Annotation and Cost-Effective Acquisition of Language Resources for Human Language Technologies"), a STREP project funded by the Language Technologies Area, Information and Communication Technologies from the 7th Framework Programme. During its 3-year life, PANACEA has developed a factory of Language Resources that automate the production, updating and maintenance of Language Resources required by Machine Translation systems and other language technologies. The factory is a web-service based platform that integrates advanced technological components for monolingual and parallel text acquisition and pre-processing, parallel corpora alignment, bilingual dictionary production, monolingual rich information lexica production, etc. (Poch et al., 2012)

The web services offered are available through the PANACEA registry¹⁷, a web-service catalogue created and maintained by ELDA. An interesting feature of the project is also the possibility to create, store and share

¹¹ <http://www.elra.info/Free-LRs,26.html>

¹² <http://www.clef-initiative.eu>

¹³ http://catalog.elra.info/product_info.php?products_id=1218

¹⁴ <http://www.promise-noe.eu>

¹⁵ <http://www.khresmoi.eu>

¹⁶ <http://www.panacea-lr.eu>

¹⁷ <http://registry.elda.org>

workflows. The already created workflows are available to the community through the workflow catalogue myPANACEAExperiment¹⁸. This catalogue has also been set up and is maintained by ELDA. Moreover, ELDA has participated in the evaluation of the platform and its different components, being responsible for the platform meta-evaluation. In addition, ELDA was responsible for the dissemination and exploitation activities, with a strong focus on IPR-issues clearing for the data used and later distributed through both the ELRA Catalogue (<http://catalog.elra.info>) and the project's web site.

PEA-TRAD is a French DGA¹⁹-financed 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. Both technology development and evaluation have involved a large data production effort, both from the point of view of the number of parallel corpora produced as well as from that of the data size produced for some of those languages. For instance, with regard to Pashto, a very low-resourced language with no associated HLT, several corpora have been produced, so as to allow for both training and evaluation:

- A 100M word monolingual text corpus has been created, which has implied identifying the sources, clearing IPR issues, collecting the data, and validating quality and content as well as formatting (Mostefa et al., 2012);
- An about 100-hour broadcast-news corpus was also produced for the Pashto language, with recordings coming from different sources so as to cover different dialects. These data have been transcribed orthographically;
- A 2M (per language) Pashto-French parallel corpus has been produced: in order to do this, 100 further hours were recorded and transcribed, so as to obtain an about 2M word Pashto corpus of transcribed speech which has been later translated into French.

The full data production effort has been carried out by ELDA, who has put into place a very strict protocol for data collection, translation, revision and quality control, in particular for all evaluation sets.

In the framework of the *QUAERO* program²⁰, ELDA has carried out the production of parallel corpora for the evaluation of Machine Translation and Speech Machine Translation involving Arabic, Chinese, English, French and German, in a number of language-pair combinations. Besides the data creation itself, this regular and large data production work (2009 through 2013) has contributed to the definition and improvement of data production protocols and

guidelines (Arranz et al., 2013). These specifications detail all the steps within the data production procedure, covering also points such as team recruitment and testing and quality control. Regarding the evaluation text data, parallel corpora contain about 27,000 words per language, while the parallel transcription corpora (from broadcast news) contain about 22,000 words per language.

4.3 Multimodal Technologies

Supporting the advance of multimodal and OCR-based document management is also another area of interest for ELRA. In this regard, ELDA has had a very significant role around evaluation for two very large French DGA-financed initiatives: the *REPERE Challenge* and *MAURDOR*.

The *REPERE Challenge* consists in evaluation campaigns for multimedia information processing systems. It aims to support research on the recognition of people in multimodal conditions. More specifically, it deals with person recognition within TV programs and in order to assess technology development, annual evaluation campaigns have been organised between 2012 and 2014²¹. As a partner involved in the evaluation part of the challenge, ELDA is responsible for the 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 both digitized and faxed documents and having a special emphasis on efficient OCR solutions. A crucial part for the R&D and evaluation tasks carried out has been the production and annotation of the necessary data: 10,000 documents (for training and testing) have been produced by ELDA, with the high complexity and challenge implied behind one such production. Such documents consist of scans of hardcopy documents which are representative of daily human operations. However, given the very sensitive nature of the documents and so as to avoid any infringement of Intellectual Property Rights, a considerable part of the corpus has been produced from scratch instead of being collected from existing material. In order to do so, the following characteristics have been taken into account: a) Multilinguality: with documents drafted in French (50%), English (25%) and Arabic (25%) by native/almost-native scribes (writers); b) The scripts are handwritten, typed characters or a mix of both, and images are also part of the documents (photographs, drawings, etc.); c) 1,553 different scribes contributed to the corpus so as to ensure enough variety. For further details on MAURDOR, their achievements and data produced, the reader can refer to (Brunessaux et al., 2014).

¹⁸ <http://myexperiment.elda.org>

¹⁹ Direction Générale de l'Armement

²⁰ <http://www.quaero.org>

²¹ <http://www.defi-repere.fr>

5. Legal Issues

5.1 ELRA & Legal Issues

ELRA has been confronted to the legal issues that prevent exchanging and sharing Language Resources from the start of its activities. It has devoted huge efforts to simplify the licensing processes while continuously monitoring the impacts of regulations changes on our community activities and trying to liaise with the various involved institutions for common actions. The first major action was to come with a few licences to allow the use of its catalogued resources by the research and the industrial sectors. Recently, its strong involvement in the META-SHARE infrastructure focused on the design and drafting of a small set of licences, inspired by its licences but also accounting for the new trends of permissive licences and free resources, represented in particular by the Creative Commons²².

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. A beta-version is planned for the summer 2014.

In addition to its role to inform the community of new developments that may have an impact on its activities, ELRA is actively involved in institutional discussions to ensure that the HLT concerns are taken into account whenever possible. For this purpose, ELRA regularly brought up the need for a copyright amendment to include an exception to the current Copyright Law in the European Union, arguing for a doctrine similar to “Fair Use” for Research purposes of copyrighted material that we see in many countries (Vienna, FlareNet Forum 2009, 12-13 Feb. 2009²³). Following on this action, ELRA contributed to a working group that debated on EC licensing for Text & Data mining and, as a consequence, contributed to an EU survey on EU copyright needed amendments.

5.2 EC licenses for Text & Data Mining

The EU organized several Working Groups on EC cross-border licensing²⁴. The WG4, in which ELRA was involved, addressed the needs, requirements, expectations, to allow Text & Data Mining on the huge knowledge databases constituted by publications assuming one has a “regular” license to view the data (as most of the libraries have). The key question was about the difference between viewing/browsing and automatic mining.

ELRA voiced its concern that technological developments and tools’ evaluations would hardly suffer from the underlying limitations and in particular that multilingual training resources would become

impossible to constitute if one needs to negotiate a corresponding license for each set and/or language. In order to allow such developments and benchmarking, the data right-holders should allow for such packaging to be shared as easily and flexibly as possible. Preventing such action would hamper the development of new technologies and hence jeopardise the activities of the stakeholders (publishers, broadcasters, etc.).

5.3 EC Copyright Survey

Following the controversial debates within WG4, the EC initiated a survey on copyright limitations in Europe that hinders/prevents activities such as Text & Data Mining but also technology development. ELRA contributed with a report on its vision about the need to substantially amend the current copyright law in Europe aiming at two major goals:

- Harmonizing the copyright regulations across the EU countries to ensure a common understanding of involved legal issues by all parties interested by local, regional and cross-border activities.
- Simplifying the access to resources by the research community, arguing for an exception similar to “fair” use for research purposes applied in many countries (e.g. USA, Korea, etc.).

The EU has plans to initiate a public debate and hopefully a change of the copyright law but no clear commitment and no schedule have been announced to this end. The contributions are publicly available²⁵.

6. Promoting Cooperation around Language Resources

The organization of events remains a key activity of the ELRA Dissemination pole. LREC (Language Resources and Evaluation Conference²⁶) continues to be our flagship dissemination event, now steadily gathering more than 1,200 experts. Some other prominent events have been held under the leadership of ELRA.

In November 2013, the NLP12 meeting gathered the 12 major NLP associations and institutions from Language Resources and Technologies, Computational Linguistics, Spoken Language Processing and Digital Humanities with the aim of boosting their co-operation and strengthening the bridges between various communities, and harmonising their activities. This meeting led to the NLP12 Paris Declaration²⁷ establishing the ISLRN.

2013 has also been the 18th anniversary of the association, which ELRA celebrated in November as well by inviting NLP experts from all over the world to participate in the workshop *ELRA international workshop on sharing Language Resources: landscape*

²² <http://creativecommons.org>

²³ <http://www.elra.info/Vienna-Declaration-of-ELRA-on.html>

²⁴ <http://ec.europa.eu/licences-for-europe-dialogue/en/content/about-site>

²⁵

http://ec.europa.eu/internal_market/consultations/2013/copyright-rules/index_en.htm

²⁶ <http://www.lrec-conf.org>

²⁷ <http://www.elra.info/Establishing-the-ISLRN.html>

and trends. To plan conferences to avoid overlaps, boost Language Resources identification and discovery, promote best practices in Language Resource citation in publications, encourage Language Resources and Tools sharing through infrastructures and strengthen relations with neighbouring communities were among the main themes discussed by the participants.

The participation in various European projects as Dissemination leaders has also led us to organize and take part in events over the last months, including the Winter School in Bressanone in February 2013, including the Promise Technology Transfer Day held at the CeBIT, in Hanover in March 2013 and the CLEF conferences (Rome 2012 and Valencia 2013), all events as part of PROMISE project, etc.

On a regular basis, ELRA is bringing financial support to a number of events: Interspeech, TALN, Tralogy, ERRARE workshop, etc.

Published by Springer, the Language Resources and Evaluation Journal, the official journal of ELRA is devoted to the acquisition, creation, annotation and use of LRs, together with methods for evaluation of resources, technologies, and applications. The usual dissemination activities continue to be carried out: monthly bulletins sent to members, web sites, etc.

7. What's next?

The next two years aim to carry on the consolidation work of ELRA. For the time being, META-SHARE mirrors several distinct repositories which deal externally with distribution issues. As a following up of the harmonization of LR Catalogues and Repositories, ELRA is working on the integration of a user management component and automatic licensing of data (inc. e-licensing) and download. This will enable a more efficient distribution with a faster access to LRs. ELRA's expectations are also focused on the ISLRN, promoting it so that it becomes the *de facto* standard of our community, ensuring the support and backing of major institutions. A second phase of the ISLRN Infrastructure setting-up aims at having the ISLRN server connected and synchronized with the major data centers' catalogues. This will allow requesters to refer to resources that are already included in existing catalogues instead of filling the ISLRN forms in detail. Of course, such actions will be supported by the ongoing struggle to catalogue more Language Resources.

Further on our consolidation work, ELRA/ELDA is involved in the recently-funded EU project MLI (Towards a MultiLingual data & services Infrastructure)²⁸. This Support Action aims to deliver the strategic vision and operational specifications needed for building a European MultiLingual data & services Infrastructure), formulate a multiannual plan for its development and deployment, and shape the multi-stakeholders alliances ensuring its long term

sustainability. For further details and discussion, we will be happy to welcome all interested LREC participants at the MLI stand that will be held at the HLT Project Village.

8. References

- Arranz Victoria, Hamon Olivier, Boudahmane Karim, Garnier-Rizet Martine (2013): *The A to Z of Manually Translated Parallel Corpora for the Evaluation of Machine Translation: Establishing Protocols through Experience*. In *Tralogy II*, January 2013, Paris, France.
- Brunessaux Sylvie, Giroux Patrick, Grilheres Bruno, Manta Mathieu, Bodin Maylis, Choukri Khalid, Galibert Olivier, Kahn Juliette (2014). *The Maurdor project - Improving automatic processing of digital documents*. In Proceedings of the 11th IAPR International Workshop on Document Analysis Systems (DAS14), April 2014, Tours, France.
- Calzolari Nicoletta, Soria Claudia, Del Gratta Riccardo, Goggi Sara, Quochi Valeria, Russo Irene, Choukri Khalid, Mariani Joseph, Piperidis Stelios, *The LREC Map of Language Resources and Technologies*, In Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC 2010), May 2010, Valletta, Malta.
- Mostefa Djamel, Choukri Khalid, Brunessaux Sylvie, Boudahmane Karim (2012). New Language Resources for the Pashto language, In Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC12), May 2012, Istanbul, Turkey.
- Park Jungyeul, Hamon Olivier, Arranz Victoria, Choukri Khalid, (2012), *Practical and Technical Aspects for Using the International Standard Language Resource Number*, In Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012), May 2012, Istanbul, Turkey.
- Poch Marc, Toral Antonio, Hamon Olivier, Quochi Valeria and Bel Núria (2012). *Towards a User-Friendly Platform for Building Language Resources based We Services*. In Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012), May 2012, Istanbul, Turkey.

²⁸ <http://www.mli-project.eu>