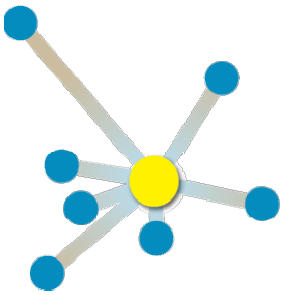
European Language Resource Coordination (ELRC) is a service contract operating under the EU's Connecting Europe Facility SMART 2014/1074 programme.



European Language Resource Coordination

Connecting Europe Facility

Deliverable D3.2.4
Task 8

ELRC Workshop Report for Spain

Author(s): Núria Bel (UPF) & Maite Melero (SESIAD)

Dissemination Level: Public

Version No.: V1

Date: 13.03.2018

Contents

1. Executive Summary

This document reports on the ELRC+ Workshop in Spain, which took place in Madrid, on January 23, 2018 at the Representation of the European Commission in Spain. It includes the agenda of the event (section 2) and briefly informs about the content of each individual, interactive and panel workshop session (sections 3 & 4).

The ELRC-Madrid Workshop was attended by 44 people. The distribution of participants was as follows: 24 people came from public administrations, 1 person from academia, 9 people from industry and 6 from other organizations.

The dedicated event page can be found at http://lr-coordination.eu/es/l2spain_agenda.

2. Workshop Agenda

2nd ELRC Workshop in Spain
Agenda - 23-01-2018

08:30 – 09:30 Registration

09:30 – 9:45 **Welcome Addresses**

Luis González
DGT – European Commission

David Pérez Fernández
SESIAD – Ministry of Energy, Tourism and Digital Agenda

Khalid Choukri
ELRC – European Language Resources Coordination, ELDA

Núria Bel
ELRC - University Pompeu Fabra

Session 1. Connecting a multilingual Europe: European context & local needs

9:45 – 10:05 **Session 1.1 Connecting public services across Europe: ambitions and results so far (VIDEO)**

Aleksandra Wesolowska
DG CONNECT – European Commission

10:05 – 10:25 **Session 1.2 National Initiatives for digital public services and (open) data**
D. David Pérez Fernández
SESIAD - Plan for the Advancement of Language Technologies
Ministry of Energy, Tourism and Digital Agenda

10:25 – 11:25 **Session 1.3 CEF in Spain: an Outlook into current and future challenges – Panel Session**
Moderator: Maite Melero
Plan for the Advancement of Language Technologies

- Enrique Maside Páramo *Director of International Relations Property,*
  Mercantile & Real Estate Registers Official College of Spain
- Ignacio Vicuña
  Director of the Centre for Judicial Documentation (CENDOJ)
  General Council of the Judiciary
- Nelson Castro Gil
  Deputy Director for Consumer Affairs,
  Spanish Agency for Consumer Affairs, Food Safety &Nutrition (AECOSAN)
- Salvador Soriano
  Open Data Coordinator
  Ministry of Industry, Energy and Digital Agenda
- Pablo de Amil,
  Head of Exploitation and Planification
  Ministry of Finance and Civil Service

11:25 – 11:45 **Session 1.4 The European Language Resource Coordination (ELRC) action**
Khalid Choukri
ELRC – ELDA

11:45 – 12:30 ***Coffee Break***

Session 2. Engage: hands-on data

12:30 – 12:50 **ELRC in Spain**
Núria Bel
ELRC - University Pompeu Fabra

12:50 – 13:10 **Can language data be shared and how? National and European legal framework**
Raquel Xalabarder
Open University of Catalonia

13:10 – 14:00 **Preparing and sharing data with the ELRC repository – and what happens next**
Victoria Arranz
ELRC - ELRA
Thierry Etchegoyhen
ELRI - Vicomtech

14:00 – 15:00 ***Lunch Break***

15:00 – 15:30 **The CEF eTranslation platform @ work (VIDEO)**
D. Markus Foti
DGT, European Commission

15:30 – 16:15 **Identifying and managing your data: Questions & Answers Mesa Redonda**
Moderator: Elena Montiel
Red de Excelencia para Recursos de Tecnologías de la Lengua – RETELE
Polytechnic University of Madrid
Victoria Arranz
Núria Bel
Raquel Xalabarder
Thierry Etchegoyhen

16:15 – 16:30 **Conclusions**
Maite Melero

16:30 – 17:00 ***Coffee Break & Networking***

3. Summary of Content of Sessions

3.1. Welcome and Workshop Objectives

Welcome speeches and the presentation of workshop objectives were the object of the first session: *Welcome Addresses*

For welcome speeches, both Mr. José María Lassalle Ruiz, Secretary of State on Information Society and Digital Agenda, and Ms. Aránzazu Beristain Ibarrola, Director of the European Commission Representation in Spain, had initially confirmed their participation. However, one week before the event both excused their participation. Luis Gonzalez from Directorate-General for Translation (DGT), was appointed as the representative of European Commission for the event and David Pérez Fernández, from the Spanish Secretary of State on Information Society and Digital Agenda (SESIAD) as the representative of the Ministry.

Mr. González opened the session by thanking everyone for their attendance on behalf of the DGT. He acknowledged the importance of languages in Europe and the need to address actions that preserve their existence. He introduced the objectives of the CEF program and highlighted the ELRC initiative as one of the actions intended to overcome language barriers. He strongly encouraged the participation of organizations and individuals to make it possible. Finally, he recalled the interest of the European Commission in Machine Translation, from the old EUROTRA and SYSTRAN times to current times, underlying the intensive use that DGT translators now make of these tools and other language resources.

David Pérez apologized for Mr. Lassalle's absence due to his official agenda. Mr. Pérez emphasized the Spanish SESIAD interest in Machine Translation and Language Technologies and how the Spanish Plan for the Advancement of Language Technology has addressed this issue motivated by Spain's linguistic richness. Machine Translation is also one of the objectives of the Plan, which will focus in the near future in improving interoperability and services for the citizens.

Khalid Choukri, from ELRC, first thanked the local organizers for the work done, and Mr. González and Mr. Pérez for the support to the workshop organization provided by their respective organizations. Then, Mr. Choukri briefly introduced the contents of the workshop.

Next, Núria Bel presented the objectives of the workshop. She emphasized that the main objective was to approach potential Language Resources, or text data, providers from different domains, and in particular for the domains where governments and public administrations provide citizens with online services. Some of these services, the CEF DSIs, were presented during the workshop. The CEF interest in Language Resources is in direct relation with the quality of Machine Translation systems, which are very dependent on the documents used to train them. Therefore, she said, it was of foremost importance that the forms provided to participants were filled in and handed over at the end of the session

3.2. Session 1.1: Connecting public services across Europe: ambition and results so far

A video presentation from Aleksandra Wesolowska (DG CONNECT) was played to the audience with live interpretation into Spanish.

3.3. Session 1.2.: National initiatives for digital public services and (open) data

To begin with, Mr. David Pérez (SESIAD) quoted the example of the World Intellectual Property Organization to emphasize the usefulness of Machine Translation for public entities involving both multilingualism and management of large quantities of data. WIPO began to use Machine Translation in 2009 and has currently developed Neural Machine Translation engines that now deliver high-quality translations of patents. In Mr. Pérez' opinion, the key factors for this successful case were twofold: on the one hand, the domain-specific data used for training (patents annotated by domain, together with their human-quality translations, were used to train the engines) and on the other hand, the pipeline of services, TAPTA, created by WIPO to manage document preparation, training and use of MT engines.

Mr. Pérez said that this example shows the benefits of integrating language datasets sources with the actual systems. He then introduced the Spanish Plan for the Advancement of Language Technology, whose objective is to foster the industry of Language Technologies by including Public Administration as the customer of these technologies. This plan is totally aligned with the CEF MT area: to develop language infrastructures, a platform to integrate services and to deploy and reuse services for specific applications. Different actions of the plan are already ongoing, and Mr. Pérez listed the agreements and contracts signed with different organizations, among which, the Real Academia Española and the EFE news agency with the aim, among other activities, to launch a platform that builds up an ecosystem for the development and use of Machine Translation, including Language Resources gathering and integration of translation engines.

3.4. Session 1.3: CEF in Spain: an outlook into current and future challenges (Panel session)

This session was chaired by Maite Melero (Officer of the Spanish Plan for the Advancement of Language Technology). Mrs. Melero prepared a number of questions in advance, suggesting the panellists to address the following points:

- Description of the digital service or services of each organization: Who is the service recipient (citizens, companies, public officials)?
- Relationship with CEF and with Europe: Does the service use any of the basic components of CEF? (eDelivery, eID, eTranslation, eInvoicing, eSignature); are there any cross-border interactions with other European national administrations?
- Multilingualism: Is the service multilingual? If so, is it currently using human or automatic translation? If the answer to the first question is negative, could it be considered?

- Potential service improvements: How could the service improve? In the sector that you represent, what digital services, which do not yet exist, are planned to be deployed or would it be advisable that they exist?

Panellists addressed these topics in the following way.

Ignacio Vicuña, Director of the CENDOJ, the Spanish National Centre for Legal Documentation. According to Mr. Vicuña presentation, the CENDOJ objectives are:

- to support judges' work, including support for the different languages of Spain;
- to provide services for the citizens,
- to deploy an infrastructure, which can also be used internationally

The most important asset of CENDOJ is its legal corpus, the world's largest legal corpus in Spanish, and the most important one for jurisprudence.

Currently they do have a jurisprudence corpus with 6,300 legal documents (trials) in Spanish, enriched with metadata and manually anonymized; a legal thesaurus of more than 20,000 terms, a bilingual corpus with translation of statements by the European Court of Human Rights, which is small but of high quality. CENDOJ has cooperated with the Real Academia Española (RAE) for the creation of the Legal Spanish Dictionary and the Pan-Hispanic Legal Dictionary.

Mr. Vicuña stated that the goal of the institution is to keep the effort of enlarging and processing these resources in the context of the Spanish Language Technology Plan. There are already plans for collaborating with IXA, a research group of the Basque Country University, to launch different Natural Language Process and Machine Translation projects in the legal domain, which will be especially interesting for offering services to citizens via the eJustice portal, which the CENDOJ also supports, so far only providing documentation.

Nelson Castro Gil, Deputy Director for the Consumer Affairs at the Spanish Agency for Consumer Affairs, Food Safety and Nutrition, AECOSAN, started by acknowledging the importance of the CEF program. His unit is basically concerned with the exchange of information between network nodes of the current authorities, although databases are open to everyone. This organization is also responsible for tasks related to the Digital Single Market and, in particular of the European Consumer Center, a platform that allows citizens to file complaints or claims, acting as a user-friendly intermediary between the citizen and the economic operator. During the coming year, many modifications are planned to improve the platform and make it more open and agile. Aecosan is also responsible for the ODR: an alternative dispute resolution platform, which is still not very well-known because of the recent directives, and little used. Mr. Castro mentioned that there are frequent complaints about the MT system that handles multilingualism.

Enrique Maside Páramo, Director of International Relations Property, Mercantile & Real Estate Registers Official College of Spain, explained that his organization is participating to the BRIS, Business Registers Interconnection System. This organization will also be participating in the Property Registrer interconnection system, which is still not operative. The European Land Registrars Association is also looking for integrating information from different State members, and Mr. Maside requested that this association be invited to future ELRC workshops.

Mr. Maside reported that nor eTranslation neither eIdentification are used in BRIS, because of interoperability issues affecting the data coming from different countries. The problem is also

legal as each country has a specific format for the information to be provided, and it is not possible to change this format without a new law. In Mr. Maside's opinion the use of Machine Translation would be very interesting, but he expresses doubts about the translation quality and the liability of the organization in case of wrong translations.

Salvador Soriano, Open Data Coordinator, Ministry of Industry, Energy and Digital Agenda explained the results of the Spanish initiatives for gathering open data from different public organizations. He reminded the audience that the actual goal of open data initiatives is to create innovative services for the citizens, to add value to public information, and he explained his view about linguistic data which, indeed, have special characteristics that make them different from numerical data. He reported on the increasing demand of text data in order to develop applications for business intelligence and opinion mining.

The Spanish open data portal has about 16,500 datasets, mostly afforded by local public administration (30% municipalities, 45% autonomous regions, and the rest from the state organizations). The portal has identified about 500 companies interested in creating innovative applications, and about 200 applications already developed.

As future actions, Mr. Soriano declared that metadata could be translated in order to support the actual use of the portal in the European Union space, as well as the connection with supercomputing facilities.

Pablo de Amil, Head of Exploitation and Planification, of the General Secretary for Digital Administration, Ministry of Finance and Civil Service, presented the Spanish Machine Translation Platform PLATA, and reported about the connection with MT@EC to support more language pairs. PLATA is currently used as an API to translate Spanish Government web pages, but is also serving other customers, such as MUFACE, Spanish Agency for Data Protection, among others. Future plans include the adaptation to the new eTranslation service as well as the deployment of a translation portal with links to external translation providers.

Ms. Maite Melero concluded the panel session by thanking the panellists for their participation and for addressing the points suggested, so as to offer the audience an overview of Spain's current status of the CEF DSIs.

3.5. Session 1.4: The CEF eTranslation platform @ work

A video presentation from Markus Foti (DGT, EC) was played to the audience with live interpretation into Spanish.

3.6. Session 2.1 The European Language Resource Coordination (ELRC) action

Khalid Choukri (ELDA) presented the action of the European Language Resource Coordination and introduced the organizations which form the consortium: Tilde, ELDA, DFKI and ILSP. Then, Mr. Choukri described the role of the National Anchor points (NAPs); in the case of Spain: Mr. David Pérez representing public administration and Ms. Núria Bel as Technical NAP.

Mr. Choukri then continued with a description of the goals of the ELRC actions: gathering language resources, identifying needs of the public sector and fostering the engagement of the public sector in the identification of language resources that the EC translation system can use to improve its engines. ELRC, he explained, is providing potential language resource providers with technical and legal support and is in fact an observatory of language resources, gathering information from workshops such as this one.
Mr. Choukri emphasized that the action's core activity was to identify data belonging to the domain of public services to improve the EC translation engines. He also reported that more than 90 resources have already been released, with more than two million translation units, which, however, was below expectations, in particular in the case of the number of resources gathered for Spanish.

Mr. Choukri ended his talk by emphasizing again the importance of eTranslation to manage the actual needs for multilingual communication and exchange of information in Europe, and the important role of sharing language resources. He then showed the ELRC web site, and the ELRC-Share repository to facilitate the access, sharing and contribution of language resources, as well the contact forms and help-desk for being assisted technical and legally to those interested in contributing with resources.

3.7. Session 2.2: ELRC activities in Spain

Ms. Núria Bel started her presentation by describing the actions that have been conducted in Spain for ELRC during 2016 and 2017, as well as an analysis of the results in comparison with the results achieved for other languages.

Ms. Bel reported that the main action carried out was the identification of possible providers, which started by sending questionnaires to the 50 identified participants at the first ELRC workshop in Madrid, in 2016. The people who filled in the forms were interviewed in person. A total of 10 persons were interviewed in 2016. On the basis of the information gathered and in collaboration with both the SESIAD and the Network of Excellence RETELE, a report on Spanish public administration as provider of resources for machine translation was released. The conclusion drafted by this report was that it was not easy to access the resources. At the same time, a survey about web site available resources was also carried out. For selecting good quality sources, a list of public organizations was elaborated including those which had issued a call at the Public Procurement Platform (Plataforma de Contratación del Estado) for translation services. Another list of potential linguistic data was elaborated out of public open data sites. For instance, TERMCAT's terminological data was identified and collected.

Ms. Bel reported that the actions carried out in Spain were similar to those carried out in other CEF countries and advised the audience to consult the results at the ELRC-SHARE repository. She pointed out that the results were not only the datasets themselves, but also the metadata needed to manage the data. She also gave some information about the size of the resources, in terms of Translation Units (TU), tokens, terms, etc., for the audience to realize how many documents are required to train a machine translation system. In order to better illustrate the point, she explained that a particular resource containing about 21,000 TU was about the same size as Don Quixote with 1000 pages, but that in order to train a system, more than 180,000 pages were needed, about the size of an encyclopaedia.

After comparing the data gathered for Spanish with the data gathered for other languages and concluding that there were not enough for improving the eTranslation engines, she described the main obstacles found by ELRC to gather data. The main obstacles were: undervalued text data, legal uncertainty and lack of data plans and protocols. Departing from these obstacles, she described the current status of language data in the Spanish public administration by locating the 10 different organizations that were studied at the previously mentioned report on a maturity model.

Finally, Ms. Bel assessed the main assets and the main problems of the ELRC actions and drafted future actions. The main positive findings were that:

1. the use of CAT tools was the critical point for facilitating the retrieval of reusable data,
2. from the list of resources collected, countries like Norway, Lithuania, Poland or Estonia have been succesful in getting different translation memories directly from Ministries and other official organizations,
3. in the countries where legal studies were carried out, it was found that most of the public administration documents can be open data.

As for findings about what is wrong, ELRC has met some resistance to modify internal document management practices, and that, in general, there is little interest in machine translation.

3.8. Session 2.3. Can language data be shared and how? National and European legal framework

Ms. Raquel Xalabarder, Professor at the Open University of Catalonia, began her presentation by introducing the legal framework, that is:

- Harmonized legal framework for EU: Directive 2003/98/CE RISP – modified by Directive 2013/37/EU
- Spanish Ley 37/2007 RISP – modified by Ley 18/2015, and Ley 9/2017 (CSP), that was a Real Decreto 1495/2011
- Spanish Ley 19/2013 de Transparencia, access to Public Information and best practices.
- Esquema Nacional de Interoperabilidad (ENI) RD 4/2010, which are guidelines and technical norms.
- The TRLPI 1996 –updated in 2017
- LOPD 15/1999 → RD 1720/2007 / RGPD 2016/679 (May'18)

Ms. Xalabarder explained that these directives and laws that encourage the reuse of public information are however restricted by other laws which, in fact, can hamper the implementation of the directive, that is, to reuse data for any purpose. Ms. Xalabarder recommends to first find out whether reusing the documents would be possible according to these other laws, which are Intellectual Property and Personal Information Protection, and work on how processing the documents would make it possible or not.

Ms. Xalabarder reminded the audience that public administration handles all type of documents. Some of these documents might contain personal information, for instance, so that they must be anonymized before any processing. Some other documents might be object of an intellectual property restriction that prevents any reuse.

In general terms she proposed to take into account the following issues:

1. Exclude any confidential information
2. Make sure that the information contained in the documents could not be linked to any physical person. This is the anonymization step.
3. Make sure that the documents, or the results after processing them, are reusable, in other words, that there is a license that allows reuse. Translations are protected works and so are translation collections and the metadata. This is important when the data is not directly generated by the administration or has not been granted with a license.
4. There are other resources which are not considered works so that they are not protected by intellectual property rights but that can have restrictions as far as reuse is concerned. Normally these are managed by use agreements (End User License Agreements).

Mrs. Xalabarder concluded by pointing out the fact that despite there are directives and laws that support the reuse of public information, there are other laws (IP and Private information) that in fact prevent its actual implementation. In addition, it should be noted that the goal of the RISP directive is to foster the creation of new online services, but the actual licenses chosen by providers are not the best fitted for these purposes, and in some cases are plainly unsuitable: for instance, Creative Commons is intended for works but some of the datasets are just data and cannot be considered works.

3.9. Session 2.4 Preparing and sharing data with the ELRC repository – and what happens next

Ms. Victoria Arranz presented details about ELRC services for gathering and sharing resources. Her presentation was complemented with new CEF actions to be carried out by the ELRI consortium.

Ms. Victoria Arranz presented the ELRC services following the ELRC provided slides. Meanwhile, Mr. Etchegoyhen introduced ELRI, European Language Resource Infrastructure, as a complementary action to boost the gathering of resources by creating national repositories under the premise that some data that could not be shared with foreign administration, could still be processed in a repository managed by a Spanish authority.

3.10. Session 2.5 Identifying and managing your data: Questions & Answers

Ms. Elena Montiel, professor at the Universidad Politécnica de Madrid, chaired this Q&A session. After a first introduction on the topic of Data Management Plan, in a kind of 'hands-on' exercise, she proposed the audience to support the action by actually filling the Engagement forms. She provided with more information about the questions in the form and invited the audience to make questions about them.

After giving some time to the audience to fill in the forms, she invited the audience to query the people at the panel: Xalabarder, Arranz, Etchegoyhen and Bel. The questions and answers are reported in section 4.3 of this report.

3.11. Session 2.6 Conclusions

Ms. Maite Melero summed up the important topics that were presented during the day. She reminded the participants that the CEF program has been funding different actions, and that the focus now are Digital Services to citizens, and that these use different common basic components, one of these being eTranslation. She then stated again the importance of language resources for making eTranslation really useful for the DSIs. In this context, the commitment of public organizations is critical because they produce lots of documents, which can in fact become open data. Ms. Melero encouraged the participants to check the data at the ELRC-SHARE repository to see whether they could have similar resources in their own organizations. She also invited them to request the assistance from the ELRC Helpdesk for any legal and technical issue they might encounter.

4. Synthesis of Workshop Discussions

4.1. ELRC and Open language data in Spain

Spain explicitly supported the PSI Directive 2003/98/CE in 2015 (in Spanish RISP). Previously, the initiative Aporta had been implemented in 2009 by the public agency RED.ES with the aim of creating the conditions for the development of the market for the reuse of public sector information, as well as to give support to public administration organizations for the publication of non-restricted access information. The current legal framework related to open data is regulated by the following legal corpus.

- Ley 37/2007 RISP – modified by Ley 18/2015, and by Ley 9/2017 (CSP) was finally compiled in RD 1495/2011
- Ley 19/2013 de Transparencia, acceso a la IP y buen gobierno (transparency, access to public information and standards of good governance)
- Esquema Nacional de Interoperabilidad (ENI) RD 4/2010 Normas técnicas, Guías, etc. (National Interoperability Scheme (ENI) RD 4/2010 Technical standards, Guides)
- Revised text of the Intellectual Property Law, 1996 – las modification in 2017
- LOPD 15/1999 about personal data processing, public liberties and fundamental rights of natural persons, and especially of their honor and personal and family privacy was revised in RD 1720/2007 and a new modification is expected in RGPD 2016/679 (Mayo'18).

A web portal http://datos.gob.es federates the open data from different public administrations: government, municipalities and regional autonomous governments. The Spanish open data portal has about 16,500 datasets, mostly afforded by local public administration (30% municipalities, 45% autonomous regions, and the rest from the state organizations). The portal has identified about 500 companies interested in creating innovative applications, and about 200 applications already developed.

There are some Language Resources in datos.gob.es (for instance, 5 datasets which are translation memories in TMX format and 1 lexicon in TBX format, all coming from the same source), but in general it is difficult to assess the number of texts available. In general, the possibility of considering texts or glossaries as open data under the PSI directive is not clearly emphasized in the open data portal related documentation and guides, many formats are allowed including some of them which refer to both numerical and text data (for instance .txt).

The top decision-maker of the Spanish data portal, Mr. Salvador Soriano, was invited to the panel (see section 3.4 of this report). As already mentioned, he emphasized the increasing demand for text data, but he did not mention any specific action towards promoting or fostering the actual collection of text data.

4.2. Success Stories and lessons learnt

Spain could not yet provide success stories for the current data collection for ELRC partly because of lack of data for the relevant language pairs. The datasets already available are mostly for translating between Spanish and co-official languages in Spain: Catalan, Basque and Galician. Besides, most institutions active in collecting data and making them accessible are autonomous regional government-dependent institutions. In the case of the central administration, some institutions, singularly the Language Interpretation Office (OIL) from the Ministry of Foreign Affairs, work with CAT tools and do have translation memories. However, legal uncertainty prevented them from sharing the resources with ELRC. There is also an unclear authorization chain for deciding what to share.

It was also found that the interest in Machine Translation from different organizations is not very strong basically because of concerns about the quality of the output. This is important also for DSIs. Two of the participants in the panel (section 3.4 of this report) referred to the quality of the translation for their respective services as a key factor.

For future ELRC activities, in what concerns Spain, the following points should be considered:

- Little use of CAT tools in the Administration, which makes the translation repositories difficult to find and handle. Facilitating the availability of these tools could make a difference.
- Direct contact with appropriate potential providers in order to better address legal and authorization concerns.

4.3. Session Questions

Due to lack of time, the only session where some questions were addressed was Session 2.5. In what follows, we sum up questions and answers in this section.

i. **Could eTranslation be used to translate a web page of a health service?** Ms. Arranz replied that eTranslation has not been trained for health-related texts and that a test should be done first.

ii. **Following question (i), in order to assess the quality and reliability of the eTranslation system a comparison with another generalist system such as Google Translate, could be used, or alternatively to retranslate in the inverse direction the test how good the quality is?** Ms. Arranz answered that using Google presents some risks because it is a generalist system whose engines have not been trained using special health data. The best validation for reliability should be human evaluation. Ms. Bel added that assessing outputs from two systems, even with a human reference, can be difficult since a system can output words different from the reference (synonyms) and still produce quality translation. Mr. Etchegoyhen added that Neural Machine Translation engines can be trained with general domain documents and that domain-specific dataset can be added at a later stage to specialize the system.

iii. As for copyright problems, the experience shows that people prefer not to perform any actions that may result in legal proceedings. Ms. Xalabarder was asked about her experience about real court actions after some undue resources reuse. Ms. Xalabarder answered that it is necessary to have an authorization for copying, exploiting, transforming a resource unless there are legal exceptions to foster data reuse. These limitations to copyright restrictions are not formalized in legislation by now, hence the permissions are granted through licenses and agreements. This means that copyright and intellectual property rights need to be refined to allow reuse in specific conditions. It is obvious that one can assess the risks of an infringement to the law and decide to undertake a "tolerated use", that is, that infringement is not likely to result in legal proceedings. This is more and more frequent, especially when there are no commercial or economic benefits. But it is still an infringement. Not to reuse data for fear of the legal consequences is detrimental to innovation, which is the spirit of the PSI directive, and in fact it results in missing opportunities. Ms. Bel said that Google is using 3rd party data for their translation system because they have probably assessed the risks. Ms. Xalabarder replied then that Google is acting under USA legislation which is different from most European laws and that they count on the "fair use" limit to copyright law. Mr. Etchegoyhen agrees and points out that USA's "fair use" is really making a difference for innovative ideas. Mr. Choukri said that there is a different point of view about the risk as the trial for Google books demonstrated. Google defended that they were not damaging the interests of book publishers because Google Books was in fact creating more demand. Furthermore, Mr. Choukri warned about new regulations in Europe, even more restrictive, and denounced how unfair it was to play on the same field as American players, but with very different rules: different legislations, less resources and limited access rights. Ms. Xalabarder points out that the USA and European laws define exploitation and transformation in a similar way, however USA has limitations to copyrights, and this is what is missing in European legislation. Ms. Montiel also participated in this discussion and came up with a question on the exploitation of a resource that was compiled from a non-authorized previous work.

Ms. Xalabarder replied that the infringement got propagated along the chain, and that ignoring the previous step could not constitute an exemption from liability.

iv. Translators are worried about their role with the emergence of Machine Translation systems and wonder what is going to happen to them. Ms. Arranz thinks that within the digital market context, the demand for translation is huge and delays are short. Machine Translation is required to complement the work of human translators. Other participants joined the discussion about the role of translators. One said that currently translators can translate more thanks to Machine Translation engines, but the work has totally changed, shifting from an almost "artistic" task to an industrial process.

5. Workshop Presentation Material

The presentations are published in both Spanish and English on the respective agenda pages at http://lr-coordination.eu/es/l2spain_agenda and http://lr-coordination.eu/l2spain_agenda