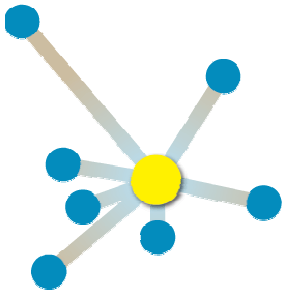
European Language Resource Coordination (ELRC) is a service contract operating under the EU's Connecting Europe Facility SMART 2014/1074 programme.



European Language Resource Coordination

Connecting Europe Facility

Deliverable Task 6

ELRC Workshop Report for France

Author(s): Hélène Mazo, Valérie Mapelli, Meritxell Fernandez-Barrera, Khalid Choukri (ELDA)

Dissemination Level: Public

Version No.: V1

Date: 17.06.2016

Contents

1. Executive Summary

This document reports on the ELRC Workshop in France, which took place in Paris, on the 11th of May 2016 at the Ministry for the Economy and Finances. It includes the agenda of the event (section 2) and briefly informs about the content of each individual, interactive and panel workshop session (sections 4 & 5). After the workshop opening and welcoming of all participants, the workshop goals were exposed, and an overview of the linguistic context in Europe was given. Then, the language technologies and public services in France were discussed.

The CEF.AT platform, which ELRC presents in the workshop as the successor of MT@EC, uses an automatic translation (AT) system based on statistical methods. This technology was explained to participants, as well as the kind of data that is needed to improve automatic translations. The legal framework for releasing data from public sector bodies was also addressed. The relevance of this AT platform for the public services in France was stressed out, and the audience was encouraged to participate in improving CEF.AT by providing the right kind of data needed by ELRC.

During the workshop, participants were informed about two ways of delivering their data: by sending them to the Consortium representatives with a copy to the helpdesk or by uploading them to the ELRC repository (http://elrc-share.ilsp.gr/).

The dedicated event webpage can be found at http://lr-coordination.eu/france.

2. Workshop Agenda

Agenda for the CEF.AT Workshop on Machine Translation

Ministères économiques et financiers
Amphithéâtre du CASC
139, rue de Bercy, 75012 Paris

08:30 – 09:05	Registration	
09:05 – 09:25	**Welcome from EC & DGT**	Isabelle Tranchant, DGT, Head of the French Dpt
09:25 – 09:40	**Overview of Connecting Europe Facility, (CEF)**	Jean-Jacques Léandri, Secretariat-General for Government Modernisation (SGMAP), French Representative at the CEF Coordination Committee
09:40 – 09:50	**Objectives and Programme of the workshop**	Khalid Choukri, ELRC Consortium, ELDA
09:50 – 10:05	**Europe and Multilingualism**	Kimmo Rossi, European Commission, DG CONNECT, Unit G3 – Data Value Chain
10:05 – 10:25	**Languages in France**	Loïc Depecker, DGLFLF Delegate General, Ministry of Culture
10:25 – 10:45	**Language Technologies in France**	Pierre Zweigenbaum, LIMSI-CNRS
10:45 – 11:15	*Coffee Break*	
11:15 – 12:05	**Panel: Multilingualism and Public Services in France**	Moderator: Loïc Depecker, DGLFLF Delegate General, Ministry of Culture Panelists: Alain Repaux, Head of the Translation Center, Ministry for the Economy & Finance Jean-Michel Thivel, General Secretariat for European Affairs (SGAE)
12:05 – 12:35	**Automated Translation: How does it work?**	François Yvon, LIMSI-CNRS
12:35 – 13:00	**How can Public institutions benefit from the CEF.AT Platform**	Kimmo Rossi, European Commission, DG CONNECT, Unit G3 – Data Value Chain
13:00 – 14:00	*Lunch at the NOVOTEL PARIS BERCY Hotel, 85, Rue de Bercy, 75012 Paris*	
14:00 – 14:20	**What data is needed by the European Commission? Practical and technical issues.**	Khalid Choukri, ELRC Consortium, ELDA
14:20 – 14:50	**Legal Framework & PSI Directive**	Danièle Bourcier, CNRS, Head of the CERSA "Law & Technologies" group
14:50 – 15:40	**Panel : Linguistic Data in France**	Moderator: Edouard Geoffrois, French National Research Agency (ANR) Intervenants: Nadia Amellah-Chikh, Légifrance Stéphane Cottin, French Government General Secretariat Brian Stacy, French National Research Agency (ANR)
15:40 – 16:10	*Coffee Break*	
16:10 – 16:45	**Interactive Session: How can we engage?**	Khalid Choukri, ELRC Consortium, ELDA
16:45 – 17:00	**Wrap-up, Onsite Conclusions, Commitments**	

3. Workshop Attendance

The event was attended by 44 participants spanning a wide range of ministries and public organisations.

4. Summary of the Content of Sessions

4.1. Opening

Khalid Choukri opens the event by welcoming the audience and introducing the key persons in conceiving and organizing the event, namely the ELRC consortium, the EC/DGT representatives and both National Anchor Points (technical and public). He then introduces Isabelle Tranchant (Head of the French language Department at the Directorate-General for Translation of the European Commission).

4.2. Welcome from EC & DGT

First, Isabelle Tranchant, Head of the French language Department at the DGT, expresses her gratitude to both NAPs (technical and public service) institutions, namely, the LIMSI-CNRS, ELDA and the French Ministry for Economy and Finances, for organizing the workshop and reminds the audience that similar workshops have already been held in almost all EU Member States. She conveys apologizes of Ms Isabelle Jegouzo, Head of the EC Representation in France, who meant to open the event but could not eventually be available today.

She recalls that May 11th comes just two days after the anniversary of the historical "Schuman Declaration", which celebrates Europe's unity.

She then gives a quick overview of the CEF programme which aims at stimulating exchanges across Europe in terms of communication, transport, energy and sheds light on the CEF.AT, one of the supporting actions for the setting up of the Single Digital Market infrastructures. CEF.AT aims at bridging the linguistic barriers among European citizens through the use of linguistic technologies, in particular MT@EC, an automated translation system with a large linguistic coverage (252 combinations) freely available to all member states public administrations.Further developments are still required to fit the translation needs of national public services. In order to train the system, adapt it to day-to-day texts, language resources are required.

Language Resources are translation memories, monolingual or multilingual corpora, lexica, dictionaries, terminological data, as well as reports, brochures, administrative documents that could serve to feed adapted corpora. The objective is to make an inventory of language data for administrations, governmental institutions for all 30 European countries. All collected data will be provided exclusively for the EC usage in the CEF.AT system.

She concludes her presentation by listing the main objectives of the workshops: raise

awareness of French administrative institutions on the available European infrastructure, allow a better understanding of the needs, discuss legal and technical aspects, encourage the preservation of the French language (the availability of multilingual corpora that include French language will reinforce the presence of the French language at the European level).

Ultimately, she asserts that demand for translation is growing, not all can be fulfilled by human translation, especially in a context of buget cuts, and that the use automated translation systems must be considered.

Isabelle Tranchant ends her presentation by wishing success to the workshop.

4.3. Overview of Connecting Europe Facility (CEF)

Khalid Choukri introduces Jean-Jacques Léandri, from the Secretariat-General for Government Modernisation (SGMAP) and also French Representative at the CEF Coordination Committee.

Jean-Jacques Leandri elaborates on the CEF programme. First, he recalls that the programme was launched prior to the Juncker Commission initiative assessing a lack of interoperability in Transport, Energy and Telecom sectors. The initial ambition for CEF was important with a 40-billion euros budget, later decreased by the European Council. For a period of 7 years, the telecommunication sector budget now reaches 1 billion euros covering the deployment of both large bandwidth networks on the European territory (150 million euros) and digital services infrastructures (850 million euro). The core platform is to be developed at the Union level under the EC authority and with a European support for the the implementation of the generic services (interfaces that dialog with the core platform) and interoperability assessement at the Member States level. Enhancing administrative cooperation with platforms for citizenship recognition, electronic signatures, invoicing, online public markets is one of the main objectives of the programme.

In the electronic administration field, an action plan was initiated some weeks ago with the following objectives:

- go digital by default, i.e. think of new services that necessarily integrate digital applications.
- offer closed systems that remain in the administrative framework.
- provide comprehensive systems for mobile users in Europe.
- offer the Domain Open Data with sets of common open metadata and data, easily retrievable to interoperate with Member States portals, like ETALAB in France.
- Set up translation process at an early stage to provide digitalized document translation in the native language of the Member State with e-delivery (electronic document exchange) in order to speed up their decisions.

He reminds the audience that the EC translation system was set up to avoid information to be translated by open systems with (sometimes) confidential data being exploited by others. He also empahizes the need for mutual enirchment, quoting the example of Norway who gave away their acquis to the EC.

4.4. Objectives and Programme of the Workshop

Khalid Choukri, ELDA, thanks Jean-Jacques Léandri and starts his presentation by providing some context. With 24 official languages, many cultures and identities, Europe is multilingual and multilingual communication is a challenge. Indeed, few people are proficient in more than one language. Translation is the way to make languages an asset rather than an issue, but in a technological world, where information is spread so quickly and so massively (he displays a map of the several throusands of tweets per day in the various European languages), human translation has to be supported by technology (automated translation – AT). Comparable situations can be found in other regions of the world: in India or in South Africa, communication in the 10 to 12 official languages triggers the same issues as in Europe.

Connecting Europe Facilities (CEF) is a program that supports the vision of a multilingual Digital Single Market (DSM). One way to accomplish this is through the new CEF.AT platform, a free platform for automated translation, provided by the DGT and based on the MT@EC translation system. This platform which will eventually allow interconnection between member states administrations and public services needs customization to improve the quality of translated outputs and be used for the benefit of all. Finding and sharing the appropriate data will help obtain the best results. Since the learning process of this system is based on human translations, better data will result in better translations. CEF.AT can help us and we can help CEF.AT and our own language by providing specific domain data.

Khalid Choukri thanks the partners involved in the organizaion of this workshop (Alain Repaux, the public National Anchor Point (Ministry for Economy and Finance), François Yvon, the technical National Anchor Point (LIMSI-CNRS), Sandrine Kerespar, Local DGT Officer.

He then clearly states the objectives of (i) sharing data, so they can be used to improve CEF.AT (ii) raising awareness and engage the participants (iii) help with legal and technical issues associated with the collection of data.

4.5. Europe and Multilingualism

Right from the start, Kimmo Rossi, DG CONNECT, Unit G3 – Data Value Chain, sheds light on the PSI directive, now transposed in the EU member states legislations, on the re-use of public sector information. He continues and describes the the European funding management in research technologies such as H2020 and CEF programmes.

He introduces CEF, the research and innovation programme in the field of language technologies, and insists on the fact that if EU claims its multilingualism, the Union has to face an enormous volume of texts to be translated.

The linguistic challenge is to automatize the tasks as human work is too voluminous. Using Google Translate is not really an option as this system does not cover all languages in addition to be rather unsafe. Within the EC, this service is authorized only for the production of published texts. The solution offered by the EC is the CEF.AT

platform that can be used by a native language speaker.

Kimmo Rossi then explains the role of each member state. CEF.AT will ensure that public digital services are available for all EU citizens, including public procurement, health services, etc., but first, it needs multilingual resources to be improved. The role of member states is to take ownership of their own language and make sure it is adequately supported in CEF-AT, by identifying and sharing the appropriate resources

On the translators' side, CEF.AT offers tools to translators. With translation volumes skyrocketing, the need for more linguist experts will grow accordingly.

4.6 Language and Language Technologies in France

4.6.1. Languages in France

Loïc Depecker, Delegate General for the French language and the languages in France (DGLFLF), French Ministry of Culture, introduces the DGLFLF, the body in charge of developing the linguistic policy for the French government.

Despite 83 inventoried languages spoken in France, France is not always acknowledged as a plurilingual country. Regional languages are part of the French plurilingualism, which requires a major linguistic planning.

He mentions Franceterme, an integrated database addressing terminology and neology issues, with the support of 20 terminology expert groups in specific domains, which contains 7,000 recommended technical and scientific terms.

The DGLFLF mission is also to promote French as a globalization language (francophony covers over 1 billion people), and the translation policy from/toFrench falls into that scope. As an example of the Delegation's interest in language technologies, a guide entitled « Mieux comprendre les outils d'aide à la traduction » (« How to understand support tools to translation ») has been published by the DGLF in 2015. DGLFLF is a linguistic research laboratory which publishes 2 book editions each month, as well as fascicles on languages of France. It also supports a number of research projects involving universities, public and private R&D labs. A project is currently ongoing on an adapted French keyboard, in cooperation with the AFNOR, the French Standardization Agency.
A call was recently launched on language technologies, with a need to identify resources available for French languages.
He expresses the wish to keep French language as a modern language and adapted to the modern world, and hopes that technologies will help French to become a globalization language, thus surviving in the digital world.

4.6.2. Language Technologies in France

Pierre Zweigenbaum, Senior Researcher at LIMSI-CNRS, replaces Joseph Mariani who could not be present. He starts his presentation by showing a graph of Language Technology field disciplines.

The main source of information on the various language technologies for French is the "French Language in the Digital Age", a White Paper part of the META-NET Language White Paper series *"Europe's Languages in the Digital Age",* addressing 30 languages and published in 2012.

Language processing is using several components at different steps: pre-processing is required before going to syntactic or semantic analysis, and consequently the full construction of sense representation. The different components are based on the use of Language Resources (LRs) as knowledge bases. In the Meta-Net White Paper for French, a table rates the status of LRs given seven criteria (quantity, availability, quality, coverage, maturity, sustainability, adaptability) and allocates a score from 0 (lowest) to 6 (highest) to each of them. For French, none of the criteria reaches the score 6.

Then, he continues with a rough overview of language technologies: spell-checking systems, machine translation, automatic summarization, text generation, speech recognition (including voice command, automatic dictation, speech transcription, emotion detection), speech dialogue systems (text generation aloud, used in station announcements for instance), text digitization, character recognition, sign language processing, web search and information retrieval. For the latter, information extraction method is used and includes entities detection, relations, negations. This technology has many applications, including anonymization work commonly used in the medical field, text classification.

As a conclusion and to stress the importance of LRs, he displays the Meta-Net Whitepaper table ranking the different Language Technologies using the same 7 criteria as for LRs. Automated translation reaches the score 5 for the (LRs) Quantity criterion but the rest of the scores are way below. Semantic processing is even less-resourced.

4.7. Machine translation: how does it work?

François Yvon, Director at LIMSI-CNRS, presents the current situation of both automatic and human translation, including computer-aided (CAT) systems. He agrees with the comments by Alain Repaux on automated translation systems current performances (previous panel) and confirms that he is ready to work on the improvement of current AT systems performance.

For now, quality translation can only be achieved through human translation. But the cost is high and the process is long.Today, translators have a whole range of online and digital tools that help them with terminology but also enable them to build translation memories (CAT tools). But the quality of translation memories remains an issue. Having more data can help improve the quality.

Then, there is automatic translation, available online, 24/7, which provides instant and low-cost/no-cost outputs, with an unpredictable level of quality that ranges from very bad to very good. However, in specific contexts, i.e. the translation of a single tweet, the automated translation systems perform well. Here again, the system is dependent on the quality and the relevance of the data that have fed the system, to improve the quality output. Automatic translation is well designed for imperfect translation and nobody is

ready to pay for its use.

In his opinion, an intermediate (or hybrid) approach combining both the translation memories and the automated translation could help. Within the Matecat project (https://www.matecat.com), the translation interface gives access to both TM and AT segments to the translator. In this regard, ergonomy is certainly an important factor.

François Yvon continues with a brief historical overview of machine translation, starting from 1955 to now, from rule-based systems to machine learning (MT is AI-complete).Statistical MT as an alternative to rule-based MT (analogy) is now a widely-adopted approach. And it requires data.

Going on with a presentation of statistical automatic translation, he provides details on the data needed and the steps to process the translation and improve the quality. On the basis of basic examples, he explains that phrase-based statistical AT approach gives better results than word-based approach, more likely to generate erroneous outputs. This phrase-based technology is implemented by Google, Microsoft, SDL-Trados, but also by Moses which is used MT@EC, the EU translation service platform.

As a conclusion, statistical AT consists of a set of tools that are used for automatic translation or that can be integrated in a human translation workflow. Qualitative data are essential and must be adapted to the different environments of institutional translation services. Monoligual corpora is needed as it improves fluidity of the translation. And of course more data is the way to obtain the most comprehensive coverage.

A question is raised by Alain Repaux at the end of François Yvon's presentation. Alain Repaux has translated the same text using MT@EC at two different periods of time and reports that the second translation output seems to be worse than the first which he does not quite understand. François Yvon answers that one option by default in some systems is that they copy-paste information. He also insists that we shall not focus on one specific result for only a small amount of text. There are evaluation campaigns that experiment and assess the systems performances (tests of robustness, silence, durability). He also asserts that statistical AT is based on a very low cumulation of calculations and variations which though may modify the frequency results.
The only way to guarantee the improvement of such systems is by adding more data.

4.8. How can Public Institutions benefit from the CEAF.AT Platform?

Kimmo Rossi, DG CONNECT, Unit G3 – Data Value Chain, thanks François Yvon for his presentation on the technical aspects of automated translation.

His opening statement is: the aim of CEF.AT is that each EU citizen should to be able to use public services online, whether or not they understand a document in a language different from their own.

Kimmo Rossi then elaborates on MT@EC. The system was funded within FP6 and FP7, developed by the DGT, and is already used by public services since June 26, 2013. The user interface was translated into 24 languages by human translators. It is a secured service. Kimmo Rossi specifies that security is not only a technical issue but also a matter

of trust. Security level also depends on the political choice of each institution. The MT@EC interface, which allows the user to either upload files or copy-paste information like in Google Translate, uses a medal (bronze, silver, gold) to show the output's quality level.

MT@EC is based on Euramis, the result of EC translators work since 1950. Over 1 billion sentences are included in the system and 2.6 millions are added on a monthly basis.

Access fees to MT@EC services have long been discussed and the "free-of-charge-service" issue has been raised at different workshops organised by the ELRC. The decision is in the member states' hands. Results from a sustainability study will be soon published providing the analysis of various funding options for CEF.AT. For the time being, registration for public institutions is free of charge.

The CEF project was financed to improve the system and the next step for the CEF.AT platform to be considered as a platform for multilingualism is to include new MT@EC functionalities, security for online services

MT@EC already gives good translations for European texts but depends a lot on language resources. If we cannot improve it, we will have to find an alternative. Indeed, volumes of texts vary a lot according to the languages and domains. Thus adapted resources are required for different domains.

Kimmo Rossi ends by mentioning the EU Open Data Portal (http://data.europa.eu/euodp/en/data) that gathers information on different national portals and which metadata was translated using MT@EC.

4.9. What data is needed by the EC? Practical and technical issues

Khalid Choukri details the data to be collected for MT: texts (translations, aligned translations, collections of comparable texts), glossaries, terminological databases, dictionaries and lists of words in one or more languages.

ELRC is looking for multi-lingual data but mono-lingual data is also of interest to ELRC, as it helps improving the fluidity of the translations. He discusses how different representation formats (e.g. *tmx, *.xliff, *.txt, *.doc, *.docx, *.odt, *.ppt) are useful to varying degrees and stressed the importance of metadata (e.g. based on Dublin core). He shows several examples of how language resources are produced from data.

Finally, he encourages the participants to contribute to the project by providing or identifying the required textual content within their organizations.

4.10. Legal Framework and PSI Directive

Daniele Bourcier, Head of the CERSA Law & Technologies group at CNRS, gives the presentation on the Legal Framework for sharing data under PSI.

Changes in this domain are frequent and therefore there is still a lot of uncertainty.

A first distinction to make is that between *open data* and *data sharing*. Whereas open

data refers to EU public policies and is thus mainly a top-down strategy, data sharing refers to policies defined by user communities. Researchers, for instance, engaged in data sharing practices before specific public policies for data sharing were established. It is thus a bottom-up strategy which relies on informal mechanisms agreed on by community members (*e.g.* principles and best practices).

Big data is another core concept to take into account. Big data come both from the private and from the public domain (*e.g.* 2013 Global Alliance for genomics and health). Today we are facing a new paradigm: from hypothesis-driven to data-driven research. The goal is to process data without knowing beforehand what will be done exactly. Legislation will have to take into account this new trend.

Data, as such, are not legally protected, unless for some reason they are covered by copyright. They have acquired economic value within the digital environment.

Over the last years there has been a terminological change: some years ago, the core concept was *access to data*, whereas nowadays regulation focuses on the *right to reuse data*. In France, for instance, the « Loi CADA » (loi n° 78-753 du 17 juillet 1978 relative à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques) focused mainly on the right of citizens to access public documents. The Ordonnance 6 juin 2005 (**Ordonnance n° 2005-650 du 6 juin 2005 relative à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques)** transposed the first PSI Directive (2003). The Loi 2015-1779 du 28 décembre 2015 (**LOI n° 2015-1779 du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public)** introduces three basic principles:

- data have to be made available in open formats;
- data have to be provided for free, although in some cases charges are possible;
- re-use: data can be re-used for commercial or non-commercial purposes, since they are considered to be re-usable assets.

Some examples are: INSPIRE data; data.bnf.fr (data are open and they follow semantic web standards).

With regard to textual data, several legal issues arise.

Unlike statistical data, textual data often have an author and therefore copyright can impose restrictions on Text and Data Mining. In France the 10th exception to copyright is being debated in order to allow Text and Data Mining. So far, however, there is no legal definition of Text and Data Mining.

Furthermore, when databases are re-used, the *sui generis* database right can also come into play.

Besides that, further regulations may apply, such as data protection ("Loi informatique et libertés"), third party copyright (*e.g.* within a consortium), statistical confidentiality or commercial confidentiality. In some cases it may be difficult to determine who the copyright holder is (the researcher, the research lab, the consortium).

In France copyright includes proprietary rights and moral rights. Proprietary rights have an expiry date, but moral rights are perpetual and inalienable.

There are also ethical issues to be taken into account. According to the COMETS (CNRS ethics committee) ethical aspects have to be taken into account for the evaluation of data processing practices (COMETS 2015).

Ethical standards come from the community rather than from regulations. They are established by members of the community (*e.g.* the community of researchers) when confronted to issues that cannot be solved by laws and regulations.

4.11. Session 12: Wrap Up

After a fruitful day of discussions, Khalid Choukri takes the time to thank all of the speakers, organisers and participants. He sums up some of the main subjects of the day, and encourages once again the participants to contribute by sharing data with ELRC.

5. Synthesis of Workshop Discussions

5.1. Panel 1: Public multilingual services in France

Moderator: Loïc Depecker, DGLFLF, Ministry of Culture.

Panellists:

- Alain Repaux, Head of the Translation Center of the Ministry for the Economy, Industry and Digital Affairs.
- Jean-Michel Thivel, General Secretariat for European Affairs (SGAE)

The moderator, Loïc Depecker, DGLFLF, Ministry of Culture, sets the objectives of the panel: to hear first-hand information from multilingual public services and provide useful information to the ELRC.

Jean-Michel Thivel starts the discussion by highlighting that he is a user of AT systems. Moreover, he is a French delegate in the EC Online Legislation group, which aims to give access to multilingual information and increase interoperability between different pieces of information. An increasing number of texts are made available in English and new interfaces are available in English too. One of tasks the group is in charge of is the N-lex website (accessible to citizens and professionals in law at http://eur-lex.europa.eu/n-lex/index_fr.htm). On this website, set up at the beginning of 2006, documents among all national legislative texts in Europe can be found. A project from the ELI (http://www.europeanlawinstitute.eu) also enables access to information on comparative law.
The SGAE helps negotiate texts that are too voluminous. Those texts must be checked which can be an issue if no AT is available.

Alain Repaux reports on his experience of automatic translation. According to him, the current AT systems can only enable an overall understanding of documents that one can read in some languages. He experienced this recently with the translation of texts (in the fiscal domain) in Latvian into French. He regrets that whatever the language is, n-oone in a French service is able to post-edit whereas this is done at the European level on an externalisation way.

5.2. Panel 2: Data and language resources for France: where to find them

Moderator: Edouard Geoffroy, French National Research Agency (ANR)

Panellists:

- Nadia Amellah-Chikh, Légifrance
- Stéphane Cottin, General Secretariat for the Government
- Brian Stacy, French National Research Agency (ANR)

Nadia Amellah-Chikh presents different corpora available on the LégiFrance website, the French Governement site for the publication of all legal information including the Frenhc Constitution, French, European and International legislation and regulations. She also mentions legal corpora from international institutions. Although some corpora were deleted on external websites, Légifrance kept copies in an Access database. She points out other public service websites that may also contain interesting data (e.g. Conseil d'Etat, Conseil constitutionnel).

M. Stéphane Cottin introduces the ECLI website (European Case Law Identifier), with several Language Resources. This includes a European repository to access multilingual data. A multilingual version of its metadata is also available.

Brian Stacy presents translation process at the ANR. The agency goes more and more international and is likely to become a bilingual agency, particularly through its website and publications. He then briefly presents the ANR, a funding agency for research projects, now employing around 280 persons, created in 2005 under the authority of the Ministry for Primary, Secondary and Higher Education and Research. Over 5,000 projects are currently ongoing with substantial budgets and ambitions.

ANR uses Trados and the TM accounts to 9611 translation units, 160K words. A terminological database of 60 terms is also available.

Different Language Resources are available at ANR including activity reports, action plan, generic call for projects, various news and press releases, legal and administrative documents, flyers, etc.

He confirms that CEF.AT aims to be integrated in the ANR schema of work and encourage everyone to mention CEF.AT.

If the translation task becomes less expensive and requires less effort, this may change the ANR vision with respect to multilingualism. This would be interesting for some bilateral partnership and would enable to translate some reports in other languages.

5.3. Interactive Session: How can we engage?

Moderator: Khalid Choukri, ELDA

First Khalid Choukri introduces the ELRC consortium website (http://lr-coordination.eu/) which provides information on the CEF.AT, the partners, the 26 workshops already organized (4 remain to be held). From this site, one can also access the online Helpdesk and ask questions on the technical and legal aspects of data sharing. Legal and technical experts support the Helpdesk operations. Questions can be asked through Skype, email and telephone. Finally, the web site offers a direct access to a repository where data can be uploaded, or else data sources can be provided.

Khalid Choukri would like to hear what the opinions and feelings of the audience are regarding ELRC and also make sure that all participants sees the importance of their data.

The discussion outlined the interest of the participants to get involved in the ELRC initiative and to contribute with different monolingual and multilingual language resources available at the institutions they represented: e.g. aligned translations of different typologies of texts; monolingual corpora, possibly enriched with linguistic annotations; domain dictionaries and terminologies. In some cases, issues about the IPR and legal status have still to be clarified. Some of the participants engaged themselves to notify and inform the heads of their departments and the higher staff at their institutions about the topics discussed at the Workshop.

5. Workshop Presentation Materials

The workshop presentations can be accessed at the event webpage, at http://www.lr-coordination.eu/france.