



Combine EU Terminology with Communication and Ontology Research

Rodolfo Maslias

► To cite this version:

Rodolfo Maslias. Combine EU Terminology with Communication and Ontology Research. Terminology and Knowledge Engineering 2014, Jun 2014, Berlin, Germany. 9 p. hal-01005876

HAL Id: hal-01005876

<https://hal.science/hal-01005876>

Submitted on 13 Jun 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Combining EU Terminology with Communication and Ontology Research

Rodolfo Maslias

Terminology Coordination Unit, European Parliament, Luxembourg

Abstract

The Terminology Coordination Unit of the European Parliament has an original approach among the Coordination Services of the 10 EU Institutions. It has opened up EU terminology to the science and industry aspects of the field.

This presentation will focus on

the access to EU terminology that TermCoord offers through a public website and free tools, through the Interinstitutional Terminology Portal EurTerm and its expert collaborators, through traineeship opportunities, study visits, open seminars, the publication of studies, university courses, terminology projects based on the interactive IATE template, participation in major international terminology networks, etc.

the constant improvement of the content and functionality of the EU database IATE; including free access to its content, cross-referencing features, as well as TermCoord's cooperation with several universities in order to explore possibilities for adding an ontological dimension to its structure of more than 100 domains.

Key words: Terminology, Cooperation, Communication, Access, IATE, Universities, Portal, Ontologies

I. Communicating for Terminology

1st Level: Internal Communication and Awareness Building

Coordinating Terminology Management in a Multilingual Public Institution

The European Union issues legislation in 24 languages and works with 552 language combinations. Among the tools used for translation on an interinstitutional level, the EU regrouped all separate databases in an interactive database permitting permanent input by the some 5000 translators of the participating Institutions. The management of this database requires a very complex system of rules and guidelines (i.e. a handbook for best terminology practices in referencing and creating definitions, legal notice addressing copyright, ownership and responsibility issues) to ensure linguistic quality and consistency between the legislative texts in all languages. In the European Parliament, which gradually obtained more and more power in the legislative process until it became the final legislator under the Lisbon Treaty, the input of terminology is created mainly from the thousands of amendments of 766 MEPs, pertaining to 7 political groups and 23 committees dealing with all policy fields. The input was organized in each translation unit in a different way and according to different approaches depending on the various usage conventions and cooperation methods of each language community and each country. In 2008, the European Parliament decided to create a horizontal unit to coordinate, harmonize, support and assist the terminology research and storage of the findings into IATE and also to cooperate with the units and the other Institutions in the huge effort of cleaning and updating a database containing millions of terms coming from different databases, massively imported without any filtering, some of them being obsolete and outdated and creating duplicates with newly inserted terms. They also needed completion with the new languages that joined the family of the EU official languages in 2003. To achieve this coordination, the new unit needed to raise awareness on the inherent role of terminology in

translation, especially of a legislation governing 28 countries and permitting each citizen to go to court on the basis of the translated version. It needed to organize a group of translators dealing with terminology in the units; today this group is the Terminology Network and numbers more than one hundred translators in the 23 language units with at least two per unit. Special skills and knowledge are needed for the "terminologists", who have to be appointed as such by their superiors among the translators to accomplish the terminology tasks.

These tasks had to be inserted into the workflow and accounting system. Training was needed, tailored for the different target groups, advanced administration of the database for the terminologists, input instructions for the translators, as well as basic training for the newcomers and for the average 60 translation unit trainees who work at the Parliament for a period of 3 months. These trainees dedicate two weeks of their traineeship to field or language specific terminology projects, both for educational and professional experience reasons and to cover needs of lacking human resources because of increasing budgetary constraints combined with an extreme workload in the translation units. Since 2012, TermCoord has been able to offer terminologists the possibility of an on-line ECQA Terminology Management Certification in cooperation with TermNet and covered by the training budget of the EP. TermCoord staff is also invited to provide training to other Directorates in the EP, interpreters and web-editors, as well as to other Institutions like to the translators of the European Investment Bank and the Court of Justice.

Different IT tools were needed such as term extractors, a macro to ease preliminary storage of terms during translation, a search engine to browse the hundreds of links to specialized glossaries imported, a webpage concentrating the links to all reference documents of all EU Institutions. Some of these resources (DocHound and GlossaryLinks) have also been made public on the website of the unit and are used also by the external contractors translating 30% of Parliament's texts.

All these efforts needed a strong communication policy based mainly on an internal website in the intranet, making part of an internal portal of the Directorate General of Translation. Awareness campaigns on different slogans like "recycle your terms", "in termino qualitas" or "tailored training" and "let's IATE, your database is what you make of it", as well as a series of terminology seminars under the generic title "Terminology in the changing World of Translation" and on topics such as "multilingual legislation process", "terminology and CAT tools", "e-lexicography and terminology in the media" that have been very well attended and evaluated by translators of the European Parliament and other Institutions.

2nd Level: Interinstitutional Communication and Information Sharing

Cooperating and Administrating a Common Interactive Database and a Terminology Portal

The internal and interactive part of IATE needs a tight and continuous cooperation between all the participating Institutions. This cooperation concerns:

- a) The content, consolidation, merging, constant updating of methods and guidelines, common or coordinated projects in specific fields; also decisions on import of collections, and referencing criteria to be followed by all EU translators
- b) Administrative aspects such as rights and roles attributed to different groups of users, transfer of ownership to another Institution for updating or completion and interinstitutional validation
- c) Technical aspects such as prioritizing improvements to functionalities made by the permanent external developers team established in the Translation Center in Luxembourg, testing and bug management.

All these decisions are taken by the IATE Management Group, accountable to an Interinstitutional Directors' body. With a presence in the Management Group since 2008, TermCoord brought about several changes increasing its efficiency by suggesting the creation of several working groups and task forces dealing with the

data entry and interface improvement, with the updating of the handbook and the integration of IATE in the new CAT tools. It also suggested the creation of an interinstitutional platform for the sharing of information and providing common resources to all EU translators, as well as the conversion of IATE in a modern web 2.0 version. Both of these initiatives have been endorsed by the interinstitutional hierarchy and a specific budget was allocated.

The EurTerm Terminology Portal developed by TermCoord in "Confluence" will be fed by an editorial board composed from the Terminology Coordination Departments and will contain access to internal IATE with which it will be connected, a centralized access to all interinstitutional and worldwide terminology resources and databanks and access to communication platforms on a language level (terminology wikis and forums).

Additionally, it will provide access to the intra-EU meta-search terminology tool QUEST searching in the most important national terminology databases in addition to IATE, the translation memory Euramis, and EurLex. Further features will include a blog, a calendar of events, a multimedia section for webinars or web streaming of events, as well as the search tool for the links to glossaries selected and updated by our Unit. TermCoord prepared in 2011 the model of this portal that has been agreed and will be finalized and opened to the translators in 2014 and in the future partly to the wider public.

TermCoord established and regularly hosts videoconferences with the Terminology Coordination Departments in Luxembourg and Brussels on strategy and management issues, deciding on communication issues, training and even an exchange of staff members for a period of two months to increase knowledge transfer and cooperation.

Except IATE, the unit also represents the EP in the management of the interinstitutional text related forum ELISE, in which translators from the three legislative Institutions translating texts related to the same act discuss on-line terminology in their respective language.

3rd Level: Cooperation with international organizations

On an international level, the Joint Inter-Agency Meeting on Computer-Assisted Translation and Terminology (JIAMCATT) regroups all international organizations and deals with all subjects related to CAT tools and Terminology. Within the scope of its main goal, which is the cooperation among all partners and the exchange of knowledge and material also in the field of terminology, these organizations try to gather all terminology resources and grant access to each other and to share their experience with terminology projects.

JIAMCATT's newly set up Terminology Working Group (consisting of representatives of EP TermCoord, UN, European Court of Justice and European Commission) contributed substantially to enrich the terminology section of JIAMCATT's website with the EU and other terminology resources. Many of these resources can be found on EU interinstitutional terminology portal EurTerm, which in the near future will be opened to all JIAMCATT members.

In the (most recent) 27th JIAMCATT meeting (April 2014), EP TermCoord proposed to create an interinstitutional access page to the members' databases and portals. The idea behind this is to offer all members of JIAMCATT a common meta-search tool (restricting access to pages containing sensitive information) that will enable searches in all databases including IATE, UNterm, FAOterm, AGORA (OECD), etc.

Sharing of terminology resources and tools among international organizations was recognized as a very beneficial exercise for the quality of translation and the need for enhanced communication and networking has been stressed in order to face the spectacular rise of the importance of terminology in a globalised context.

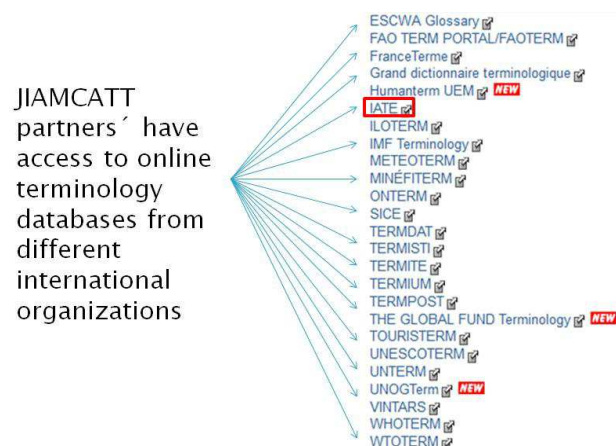


Fig. 1 Exchanging know-how and material in the field of terminology

The common platform will provide a single central access to terminology resources, tools and databases; efficient and user-friendly interinstitutional communication, linguistic knowledge and skills sharing as well as enabling communication on language level and on specific topics.

4th Level: External Communication, Promotion and Coining of Resources

Connecting with the External World of Terminology

Terminology has evolved in the last years, thanks to and because of the inherent need for multilingual communication for every public, private and academic sphere of activity in a globalised world.

Important companies worldwide possess huge terminology departments and databases; private providers or non-profit organizations offer databanks and terminology support. Also, very large professional forums on terminology have been created, each international activity includes a glossary on the relevant website, there are yearly hundreds of conferences on terminology and connected activities, very large organizations regroup hundreds of public, private and academic actors dealing with terminology.

Terminology has also become a separate discipline in the linguistic sector and the humanities. Hundreds of Universities worldwide have created terminology departments or integrated terminology courses and practice into their programs in translation and linguistics faculties. Terminology in several fields is very frequently the subject of academic theses of all levels and the Terminology Coordination Unit of the European Parliament has to select every three months among more than 1500 applications from all over the EU to cover the 6 posts for a specialized traineeship in terminology. While in the recent future, terminology specialists tried to enrich the web with terminology resources, today in the era of web 3.0 and big data, the main task is to regroup in an easily accessible way all the huge and multiple resources in all languages of the world and all fields of activity.

The European Union has the largest and most multilingual human translation production in the world and cannot stay isolated in this wonderful evolution of terminology and its adaption to the current needs of our localized and globalized world. TermCoord has started an intensive campaign and could include IATE in some of the biggest terminology search tools and webpages.

TermCoord has contributed a module on Terminology to the University of Luxembourg Master's program "Multilingual Learning in a Multicultural Environment". This consists of courses on multilingualism, terminology management, digital terminology coining and practical exercises, accompanied by theoretical courses given by Prof. Rute Costa in the next semester. The students of this module are offered internships in TermCoord. Staff members of the Unit are invited to several universities for seminars and to initiate IATE based terminology projects.

As from its creation, TermCoord has developed a wide network of external contacts mainly in the other international institutions and the academic world. Even in the first months of its existence it invited 10 prominent terminology professors and cooperated with them for the development of an original approach of terminology management in a public institution. A very important asset for the unit has been the knowledge and enthusiasm of the by now some 100 trainees from all countries with studies or specific interest in terminology, who offered an important input from the most recent developments and keep cooperating with the unit in several projects through specific accounts and discussions in the social media.

In April 2011, the unit went public with a website that attracted up until December 2013 300.000 visitors and became an important reference in the world of terminology offering EU specific resources and interesting posts on translation and terminology prepared by our staff in cooperation with trainees specializing in communication. In 2012, this external presence was extended with accounts and pages in the most important professional social media networks.

The biggest interest was shown for our pages containing the IATE term of the week, related and linking to an article on an important current event, the interviews with prominent terminologists conducted by our trainees, and the specific EU terminology resources and the different possibilities offered by the European Parliament for traineeships and study visits in our unit.

In 2012, TermCoord launched IATE projects with Universities: Interested translation and terminology departments receive a template to make exercises with their students simulating input into the database and compile lists of entries in their languages in cooperation with our unit, which after validation are inserted in IATE. These projects are still limited in the pilot phase and are intended to become in the future, projects conducted by the Terminology Coordination Departments of all Institutions cooperating in IATE.

As from the first semester of 2013, the University of Luxembourg inserted in the syllabus of its master "multilingualism in a multicultural environment" a series of courses and workshops on terminology management, multilingualism, terminology coining and web monitoring, given in three languages by members of TermCoord staff with the agreement of the European Parliament. Several universities invited the lecturers to give a part of these courses combined with a terminology project (Saarbrücken, Vienna, Thessaloniki, Bologna, Paris, Sofia, Salerno, Malaga, Vigo a.o.).

TermCoord is preparing a project to insert terminology research and ontologies in the European Master of Translation offered by the European Commission in cooperation with several Universities.

Our unit was the subject of a master thesis at the University of Vienna and provides the possibility to students and researchers to publish on its website their theses and studies on terminology or containing a chapter on terminology. Students working on terminology in their universities are allowed to apply for an unpaid traineeship in TermCoord, where they are familiarized with practical terminology work.

Several university professors were invited to our big seminars as speakers or were invited to run specific workshops for the translators and terminologists. The next seminar will take place on the 5th of June under the title "Terminology in Academy", focusing on this cooperation and ontologies.

TermCoord is member of EAFT (European Association for Terminology) and TermNet and participates in international terminology conferences autonomously or representing all EU Institutions like in the latest JIAMCATT hosted by the European Parliament in Luxembourg in 2012. Presentations are also given in Universities upon request or organized in the European Parliament in Luxembourg for groups of professors and students.

II. Extending Terminology to New Horizons: TermCoord Experiments with the Semantic Web Technologies

The discipline of Terminology is in constant evolution. It is widely considered as a stand-alone science and has been put into the spotlight by universities, companies, state bodies and international institutions worldwide. In the era of open linked data, big data and web 3.0, a highly specific subject like terminology cannot ignore neither new technology inputs from these fields nor the mutual benefits to be obtained.

In this sense, IATE Public has recently undergone several reviews and improvements. The most interesting ones include entry-to-entry links and cross-references as active hyperlinks. Moreover, a web service that permits queries has also been implemented. In particular, these changes will soon make it possible to download IATE data exports in TBX format after updating its restrictive "legal notice".

In addition, LTAC Global (www.ltacglobal.org), the host organization of the TBX Steering Committee, will cooperate with IATE in the design of a term base template that matches the content of document-specific glossaries that could then be imported into different CAT tools. With this goal in mind, the Committee will define a dialect of TBX that supports the selected fields and will provide working software in Perl or some other programming language that automatically converts between TBX and the term base format. Together with the IATE software developers, the Committee will then create an application that formats information from IATE in the selected TBX dialect.

On top of that, TermCoord has also suggested transforming IATE (linguistic resource in form of a terminological relational database) into a knowledge management resource in the form of linked data, in order to produce multilingual solutions for different working domains that are homogeneous and responsive.

The main difference between a terminological database and a knowledge base is that the former presents concepts, terms and their multilingual equivalents in isolation with no semantic interconnectivity. The latter, on the other hand, presents concepts as 3D networks with several different aspects from which you can request and obtain linguistic and cognitive information in a well-structured, logical manner.

For this reason, the IATE Management Group is looking to achieve the amelioration of IATE with the semantic web technologies. By developing a model of linguistic-terminological knowledge management, the Group wishes to define new typology for a terminologist profile within the EU Institutions. This profile is set to deal not only with concepts and terms, but also with knowledge management, multilingual webpage content indexing, and website content management.

In this context, the IATE Management Group, represented by 3 of its members (EP TermCoord, the Translation Centre and the Publications Office of the EU) will launch a project concerning the development of an ontological structure in 2 sub-domains of IATE. This will be done in cooperation with the University of Luxembourg (philosophy and linguistics), the University of Salerno (communication) and the University Sapienza of Rome (computational linguistics). This project should refer to the initiatives related to the ISA Program, such as a common metadata vocabulary or Asset Description Metadata Schema (ADMS).

The first step will be to find an interoperability solution for IATE in order to overcome the semantic heterogeneity within the Institutions of the EU, i.e. Council of the EU, European Commission, European Parliament, Court of Justice of the EU, European Court of Auditors, European Central Bank, European Investment Bank, European Economic and Social Committee, Committee of the Region and Translation Centre.

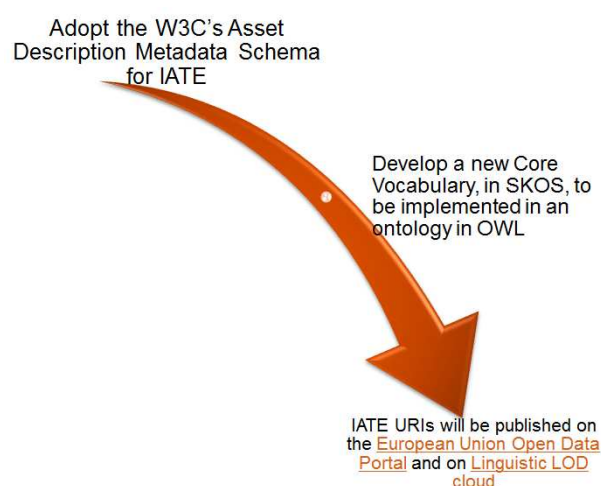


Fig. 2 Interoperability solution for IATE

On the one hand, this common project aims at enhancing stakeholders' access to EU policies, as well as supporting EU citizens in understanding institutional terminology. Yet on the other hand, it should improve the management of institutional terminology and communication and, above all, it should help translators in their daily tasks with non-ambiguous terms by providing direct links to reference texts.

The domain study will be conducted taking into account sociolinguistic, economic, organisational and translation aspects. This is intended to narrow the initial conceptualization of the field of expertise and to establish an initial conceptual structure prototype.

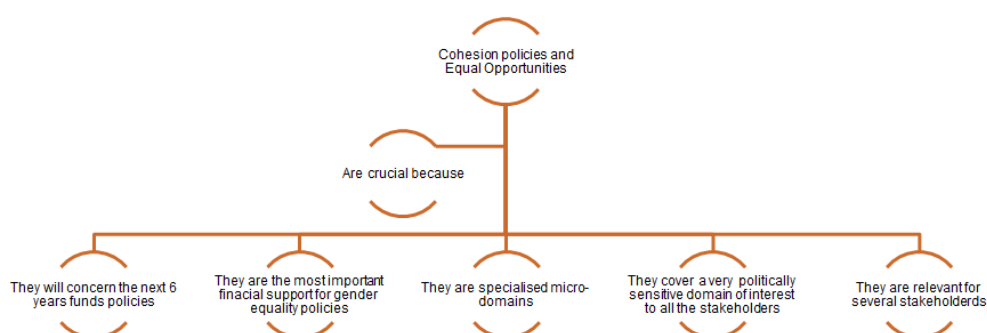


Fig. 3 Domain analysis

As a second step, a specialised corpus will be compiled to develop the database, i.e. linguistic (terms, contexts), semantic (semantic relations, concepts) and pragmatic information (frequency of use, communicative situation).

Since an extraction from a *corpus unicum* can lead to ambiguity due to the varying concepts a single expression can refer to in different contexts, a text classification will be implemented to avoid this situation.

Moreover, this facilitates linking a term to the text where it was extracted from in the ontology and in turn will improve the references and quality of the extracted terminology.

Terminology extraction (with a special focus on multi-word expressions) will be carried out following this corpus-based approach by a classification on text types for different pragmatic finalities.

The focus of this project lies on multi-word expressions since they are rarely, and often poorly, recognized by statistical extraction tools. Most terminology extraction tools that focus on frequency do not recognize multi

word expressions, because they are based on mathematical-statistical algorithms. For this reason, TermCoord will continue testing different extraction tools -from statistical, linguistic to hybrid ones- to approach this project in the most appropriate way.

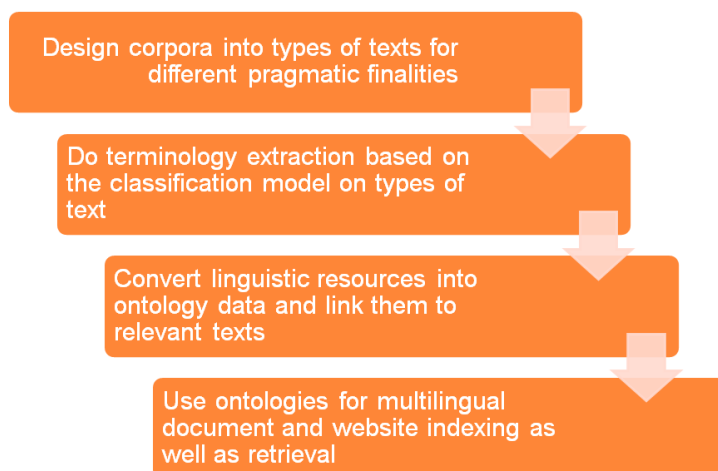


Fig. 4 The different phases of the project

To acquire knowledge from corpora, the most widely used method is based on knowledge patterns. These are logical hierarchical relationships, ontological hierarchical relationships and non-hierarchical or associative relationships. This data will be extracted from the compiled corpora in order to gain vertical, as well as horizontal relationships that will then be converted into computational linguistic resources to be used in ontologies.

Although there are different ways to develop domain ontology, either by using a linear approach or by means of graphics or semantic frameworks, if the latter is applied to terminology, it allows a clear and orderly structuring of concepts; thus, it allows observing the context in which different concepts and terms appear. Therefore, semantic frameworks will be implemented in our project and every ontology entry will be linked to a specific type of text.

Apart from the above described goals, the project also intends to examine the following aspects:

- The integration of IATE with EuroVoc and VocBench (SKOS) and its use in document and website indexing

EuroVoc is a multilingual thesaurus which was originally built up specifically for processing the documentary information of the EU institutions and now it is a controlled set of vocabulary which can be also used outside the EU institutions. The aim of the thesaurus is to provide the information management and dissemination services with a coherent indexing tool for the effective management of their documentary resources and to enable users to carry out documentary searches using controlled vocabulary.

VocBench is a common platform offered by the Publications Office to the EU Institutions and agencies for the maintenance/dissemination of controlled vocabularies as well as indexing/annotation of their documents.

- How to reduce redundancy and overlapping in EU terminology in order to foster a more coherent and standardised use of terminology in the European Union
- The use of the European Open Data Portal to share the resources (in TBX or TBXmin)

Cooperation of TermCoord with the Laboratory of Terminology of the University of Bologna

In the recent seminar, Terminology in Academia (termcoord.eu/terminology-academia/) organised at the European Parliament, Prof. Franco Bertaccini presented the difference between term databases and knowledge bases to the translators and terminologists of the EU Institutions.

In the context of this cooperation, a student of this University will dedicate her thesis to a sample IATE database including ontologies.

References

Colomb, Robert M. (2007): *Ontology and the Semantic Web*. Netherlands: IOS Press.

Elia A; Monteleone M; Marano F; Vellutino D. (2009). Strutture sintagmatiche e semantiche per il riconoscimento di polirematiche terminologiche. In: Word classes - Nature, typology, computational representation - Actes of SecondTRIPLE International Conference Roma Università degli Studi 24-26 March 2010.

Grigonyte, Gintare (2010): *Building and Evaluating Domain Ontologies: NLP Contributions*. Berlin: Logos Verlag.

McGuinness, D. L. (2002): *Ontologies Come of Age*. In D. Fensel, J. Hendler, H. Lieberman, & W. Wahlster (Eds.). *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*. Cambridge, Mass: MIT Press.

Melby, Alan K. (2012): *Terminology in the age of multilingual corpora*. In *The Journal of Specialised Translation*, issue 18.

Publications Office of the European Union: http://publications.europa.eu/index_en.htm

W3C and the Semantic Web: <http://www.w3.org/2001/sw/>

You can find any additional information on TermCoord's website (<http://termcoord.eu>)

Presentation by Rodolfo Maslias (Head of the Terminology Coordination Unit of the European Parliament)

In collaboration with: Carolina Dunaevsky (University of Cologne) and Marie-José Palos Caravina (Translation Centre of the EU)

6.6.2014