

## CHAPTER TWO

# CONVERTING THE EUROPEAN TERMINOLOGY DATABASE IATE INTO THE WORLD'S LARGEST MULTILINGUAL DATA SPACE

RODOLFO MASLIAS

The European Union provides 80 % of the national legislation of its 28 Member States, in 24 official languages. This legislation is debated and finalized mainly by the three legislative institutions, the European Commission, the Council of the European Union and the European Parliament, but numerous other institutions and bodies participate in the governance of this large economic and political confederation of states. The fact that every legislative document produced by the EU institutions can become a binding legislative act for each Member State in its official language(s) means that the multilingual aspect of the procedure is obligatory. Indeed, multilingualism was the object of the very first regulation issued by what would become the European Union, in 1958. This regulation has always kept its number as the first legislative act, but between 1958 and the present the number of official languages has risen from 6 to 24. The fact that a translated legislative act may be made use of in any Member State and by any citizen or company for legal purposes means that every translation of a legislative text has the status of an original, and this explains the need for the accuracy and consistency of all translations in all official languages. To achieve this goal, the European Union has the largest linguistic apparatus in the world, with some 5 000 translators in more than 10 institutions, along with hundreds of interpreters, translating and interpreting in 552 language combinations. The production of legislation in every field of activity and for so many Member States with different needs and structures involves hundreds of meetings needing interpretation and thousands of documents needing translation: all this is a part of everyday life in the European institutions. Thus, for example, the European Parliament's translation page count for a

single year stands at 223,000. The consistency of legislation in the different Member States and in all languages depends on the use of accurate terminology in every field of action and in every language.

This terminology needs to be stored and constantly updated, so that it can be shared by the translators of all institutions, bodies and agencies participating in the drafting of European legislation. For this purpose, the EU has created a huge interactive terminology database that contains today more than 11 million terms in more than hundred domains and in the 24 official EU languages. This database is called IATE (Interactive Terminology for Europe, <http://iate.europa.eu/>). It is interactive in the sense that each translator in any of the institutions has to consult the existing terminology of the database, but is also enabled, even encouraged, to insert into it the result of his or her research into any “problem term” encountered during translation. In this way, the database is enriched every day with an average of 600 new terms: in 2014, for example, IATE acquired 225 000 new entries. However, apart from creating new entries for new concepts and adding new terms, the interactive function of the European terminology database permits the constant updating of the existing entries, thus maintaining its very high level of reliability. The interactivity of the database is limited to the internal version, since the terms are inserted into the public version after a process of validation. These can be commented on, but not modified, by the external users of IATE. The number of these external users is impressive, since we may count an average of 3 600 clicks per hour worldwide. IATE is recognized as the most precious resource for freelance translators, being the biggest multilingual database covering all fields of legislation.

The quality of entries is guaranteed by the implementation of very strict rules contained in a very detailed handbook, based on the international ISO standards on terminology and adapted to the specific needs of a multilingual database covering legislative texts in multiple domains and addressed to numerous layers of users ranging from specialized bodies to individual citizens. This procedure includes a form of cooperation among the institutions known as “consolidation”, which permits common work on terminology projects related to the legislative procedures involving them all, the merging of entries with similar content, and the interinstitutional updating of the various fields, such as definition, reference and missing languages. The quality level of an entry is “marked” on a points scale from 1 to 4, depending on the reliability of the source and on how recent the update is. There is a final stage called validation, carried out by the translators-terminologists in the EU institutions. This action is

very important, especially for the wider public, because “clicking on the validation button” automatically exports the entry to IATE Public.

The structures of this cooperation between such a large number of institutions are considered to constitute an exemplary procedure. The cooperation is governed by a working group which is responsible to the hierarchies of the translation services in all the institutions. The IATE Management Group takes all policy decisions for the database, on the basis of the work and conclusions of numerous taskforces dealing with the different issues.

Since 2013, IATE cooperation has been enhanced by the use of a terminology portal which is accessible to all EU officials in all the institutions. This portal, called EurTerm, enables the sharing of internal and external resources, collaborative work on terminology projects, and updates with the latest information on terminology both inside and outside the EU. Furthermore, it gives central access to wikis for each EU official language, through which the language community can develop common terminology projects, discuss various topics via domain- or text-related forums, and, last but not least, involve national experts by giving them read access to the EurTerm language wiki. This platform is managed by a web editor specialized in terminology, coordinating the interinstitutional work for this portal from the Terminology Coordination Unit of the European Parliament.

Terminology is thus an important administrative activity for a large institutional complex like the European Union, but it is also much more than that: in our times of intensive global communication in every field, multilingualism becomes a great challenge and an unavoidable necessity. Communication and understanding of the same concept in all languages used for a multinational or globalised activity is crucial, and this means consistent terminology. New technologies and intensive research in new fields daily create large amounts of new terminology in all domains and in all languages. Digital progress provides support in the form of huge interactive databases and metasearch tools to store and consult multilingual terminology. Today there is no large institution or company without a specialized database or glossary to enable understanding between its partners in different countries.

This international multilingual communication and cooperation creates huge needs for translation and interpretation. Terminology is the milestone of the quality of these two activities. Technology provides more and more tools and software to ease this work: computer-assisted translation tools, and—gradually—automated translation. The research into the development and improvement of these tools is based on two supports: the corpora of

translated texts (known as “translation memories”), and terminology databases. While translation memories can eternalize mistakes, the terminology database guarantees the quality of translations and their consistency in all languages—elements that are crucial for the correct functioning of any multilingual institution or company. As with the need for consistency of European legislation as explained above, imagine how important it is that the instruction manual for a complex machine, or a scientific report in a field like medicine, should have an identical meaning in all languages!

New language technologies need specialized terminology. They also enable and ease terminology work, its sharing and storing, and cooperation between terminology producers and managers. Interlinking possibilities, metasearch tools permitting easy searches, managing of big data, cloud technologies, statistical term extraction and multilingual alignment: all these possibilities now mean that overlapping can be avoided in terminology work. Just think of the number of glossaries or databases you encounter that contain the same entry for a specialized but commonly used financial term, for instance “tax fraud”. Translators and terminologists from every financial or political institution, every bank, every university department dealing with finance or politics have created an entry in their terminology database or glossary, and in all the world’s languages.

Interlinking European terminology with all reliable specialized terminology databases managed by all international organizations and bodies, academia and industry, through clouds or metasearch technology, is an objective pursued by the European Parliament’s Terminology Coordination Unit in recent years, via cooperation and networking with the main actors in the field of terminology. This networking through the public website [termcoord.eu](http://termcoord.eu), as well as intensive activity on social media, have made it possible to launch a project involving various European universities and language service providers, which has the dual aim of converting the huge collection of selected terminology resources (GlossaryLinks, <http://termcoord.eu/glossarylinks/>) provided by the European Parliament into a powerful and user-friendly metasearch tool and connecting it with the IATE European database, and applying ontologies to the database that will make it the largest multilingual semantic terminology database worldwide. We have called this project the “Big Multilingual Terminological Data Space”, and presented it in 2015 at the Riga summit on the “Single Digital Multilingual Market”.

The very rich terminology material gathered for so many years from the huge legislative corpora of European legislation in so many languages must now respond to the challenge of sharing, interlinking and structuring

data using the most efficient communication end-sharing technologies. It will, of course, always be based on the strictest international quality standards, in order to offer, to both European citizens and to all professionals in translation and interpretation around the world, this precious linguistic asset that has been created and is constantly enriched by the world's biggest human translation apparatus: the translation and interpretation services of the European institutions.