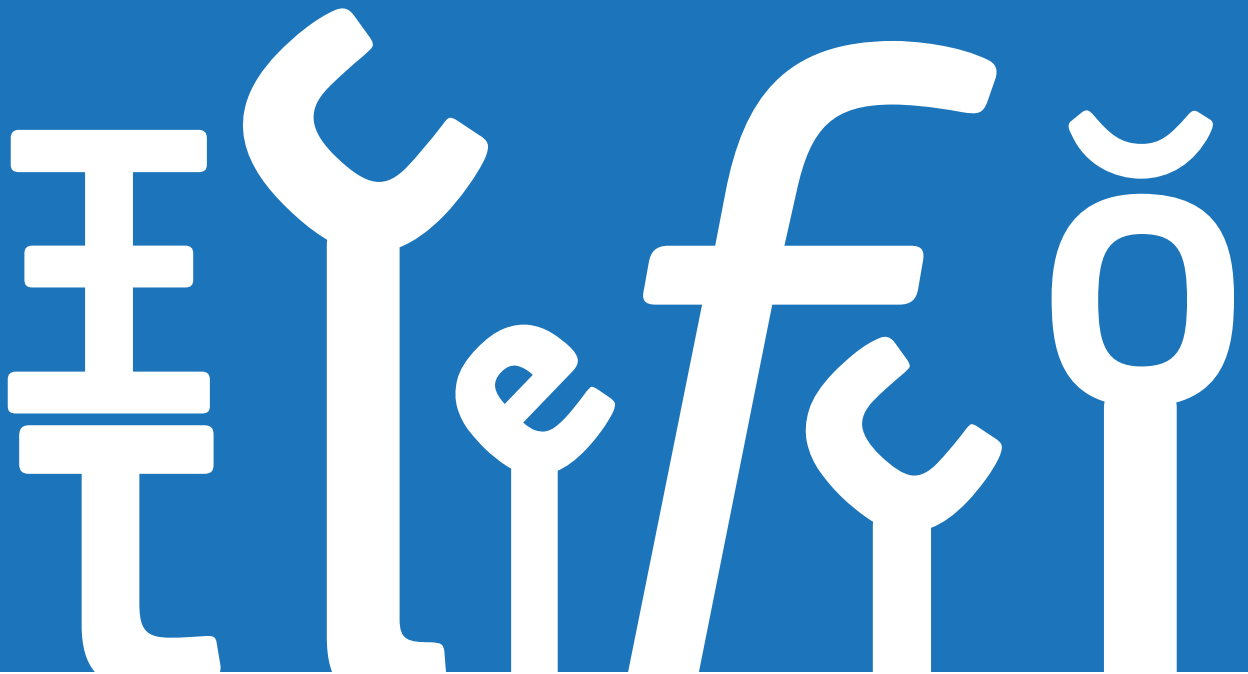




Translation tools

and workflow



Contents

Translating for the European Commission	1
Translators' needs	3
Our language resources	3
Translation workflow in the European Commission	4
Tools	6
■ Administration and documentation tools	6
■ Translation tools	9
Euramis	14
What's next?	19

Translating for the European Commission

The Directorate-General for Translation is one of the biggest translation services in the world. It is also the largest single department in the European Commission.

We have a total number of staff of around 2 500, including translators and staff performing tasks supporting translation work (management, assistants, communication, information technology, training, etc.). This figure includes all staff in Brussels, Luxembourg and the local offices in the Member States.

We translate some 2 million pages a year.

Other EU institutions and bodies (the Council, Parliament, Court of Justice, European Economic and Social Committee and Committee of the Regions, Court of Auditors, etc.) have their own translation departments, whereas

the various specialised decentralised EU agencies and bodies send their translation work to the Translation Centre for the Bodies of the European Union.

DG Translation is organised according to languages. Each official EU language has its own language department which is organised into translation units. Translators therefore work in single-language units which specialise in particular subjects. They translate out of several languages, but almost always into their mother tongue. Requests for translations go through a central demand management unit that negotiates the deadlines and how many language versions are needed of each incoming text, and then assigns the jobs to the appropriate units.





Translators' needs

Information technology is playing an ever increasing role in translators' daily work. We therefore make various computer tools available to translators, who use them according to their translation needs and personal preferences. The main document formats we work with are Word, Excel, PowerPoint, HTML and XML.

Irrespective of their preferred working methods, all translators' needs are basically the same:

- appropriate terminology (dictionaries, glossaries, terminology databases, etc.);
- reference documents (paper and electronic archives, aligned texts, etc.);
- possibility to reuse previously translated texts (translation memories, electronic archives, etc.);
- central and local assistance. The role of secretaries has evolved from being a typist into a fully-fledged translation assistant. Secretaries and translators work hand in hand more than ever, with pre- and post-processing being handled by secretaries and translators focusing on the actual translation work. Assistance is provided centrally by a helpdesk and an alignment and pre-processing team, and locally within the language units themselves.

Our language resources

To perform our tasks, we have a wide variety of language resources at our disposal:

- terminology in many different forms (multilingual libraries, terminology databases, electronic dictionaries, etc.);
- translation memories enabling genuine data sharing;
- previous translations available from internal archives and other sources;
- machine translation, which, at the European Commission, is used as a genuine translation aid by translators and as a way for requesting Commission departments to find out the gist of a text.

Translation workflow in the European Commission

Brief explanation of tools

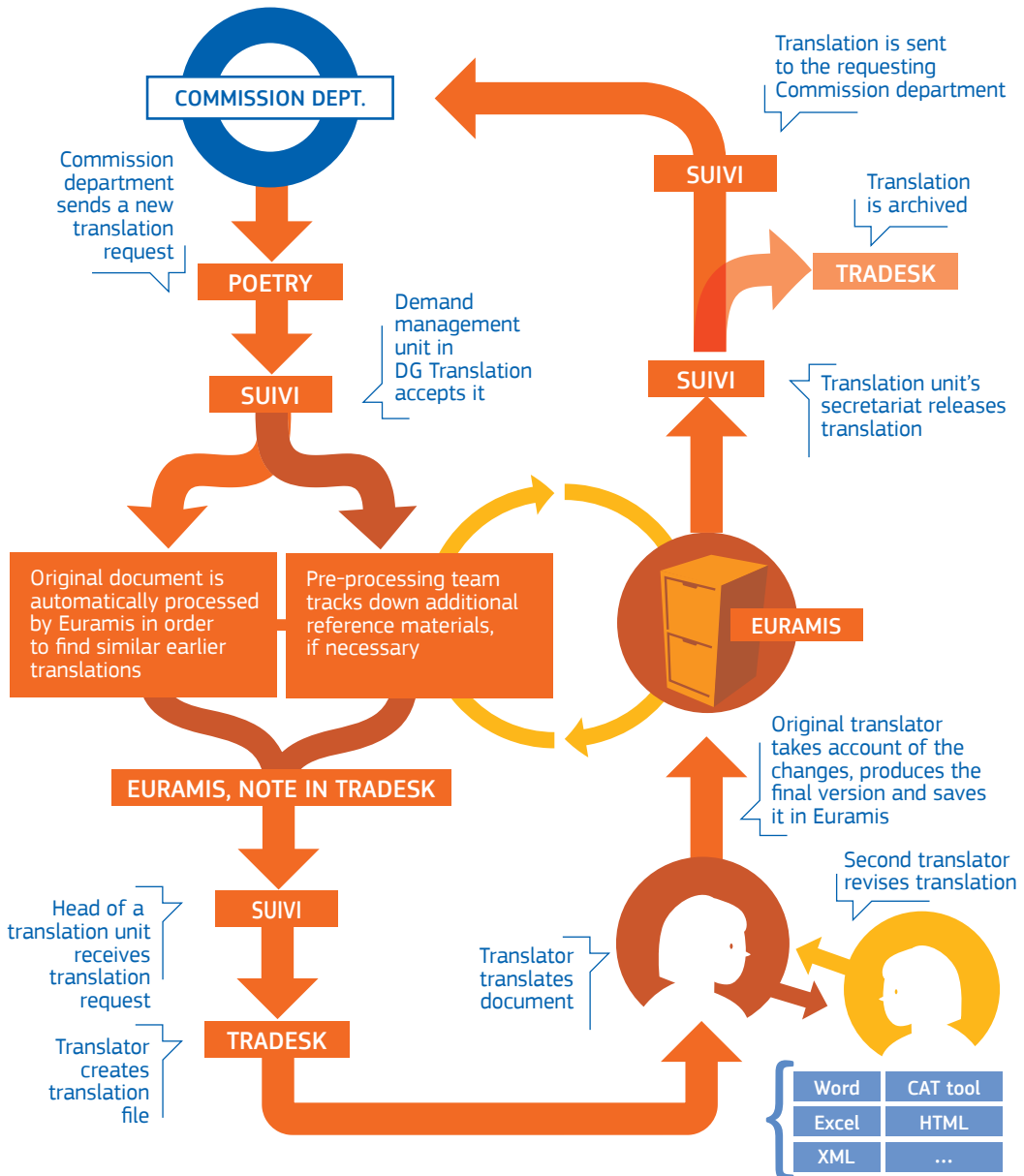
- **Poetry**
Software for electronically sending translation requests from the Commission's other departments to DG Translation.
- **Suivi**
Software for electronically managing translation requests inside DG Translation.
- **Euramis**
Central translation memory of DG Translation.
- **TraDesk**
Interface for translation document management and access to electronic archives where earlier translations are stored.
- **CAT tool (Computer-Aided Translation)**
Translation memory software for local management of already translated text retrieved from the central memory.
- **DGTVista**
Document search and view engine.
- **EUR-Lex**
Online database of European Union law (<http://eur-lex.europa.eu>).
- **IATE**
Terminology database of all European Union institutions (<http://iate.europa.eu>).

All translation requests are sent by Poetry/Suivi to Euramis for automatic retrieval of previously translated documents which are relevant for the request at hand. Results are automatically stored by Euramis in a document server. Translators have access to these pre-processed files through TraDesk.

A macro allows users to automatically create a local translation memory containing relevant translation information and fill in the metadata.

After interactive translation with the CAT tool, another macro allows users to automatically clean up, export and save to Euramis all translated documents.

TRANSLATION WORKFLOW IN THE EUROPEAN COMMISSION (*)



(*) All tools mentioned in the diagram are explained in detail in the following pages.

f Tools

Administration and documentation tools

Poetry

Poetry is the software used for the electronic transmission of translation requests from our clients (other Commission departments) to DG Translation. The web interface constructs an electronic folder containing the translation request, the original document for translation and any reference documents, all of which the requester can send for translation in one go.

Some of the many clear advantages of Poetry are:

- fast transmission;
- integration into our electronic file archiving system, TraDesk;
- availability of original and reference documents in electronic form;
- improved electronic workflow.

The screenshot displays the Poetry web interface, which is a web-based form for submitting translation requests. The interface is organized into several sections:

- General information (step 1):** This section contains fields for 'Project name', 'Status', 'Date of request', 'Date received', 'File', 'Type of translation', 'Description', 'Precedence/Reference', 'Start date', 'End date', 'Requester deadline', 'Comments', 'Service level/Agreement', 'Priority', 'Comments', 'SIC', 'Intermark', 'Phone', 'Fax', and 'Comments'.
- Original document (step 2):** This section includes fields for 'File', 'Format', 'Source language', and 'Destination language'.
- Contacts (step 3):** This section includes fields for 'Requesting service', 'Originating service', 'Contact person', 'Workdays', 'Requester deadline', and 'Comments'.
- Target languages (step 4):** This section is a table with columns for 'Language', 'Requester deadline', 'Deadline set', 'Format', 'Specific return address', 'Comments', and 'Status'.

The table in the 'Target languages' section contains the following data:

Language	Requester deadline	Deadline set	Format	Specific return address	Comments	Status
EN	18/03/2012 10:00	18/03/2012 10:00	DOC	EN: ENCC@DGTRAN.eu		Success
FR	18/03/2012 10:00	18/03/2012 10:00	DOC	FR: FRCC@DGTRAN.eu		Success
DE	18/03/2012 10:00	18/03/2012 10:00	DOC	DE: DECC@DGTRAN.eu		Success
IT	18/03/2012 10:00	18/03/2012 10:00	DOC	IT: ITCC@DGTRAN.eu		Success

Suivi

Suivi is the software used to electronically manage translation requests inside DG Translation. The system is used

to assign translations to units and to translators and to deliver translations to requesters.

TraDesk (Translator's Desktop)

TraDesk is the interface that translators use to view their current tasks and related information, e.g. job sheets, and to access their documents.

TraDesk provides access to the original document and to all files needed during the translation process, including:

- reference documents;
- pre-processing files;
- comparisons between different versions;
- ongoing translations;
- completed translations.

TraDesk is the tool for managing translations, and it includes an alert and note function to enable translators working in different

EU institutions on the same translation project to communicate with each other. TraDesk is also our electronic archiving system.

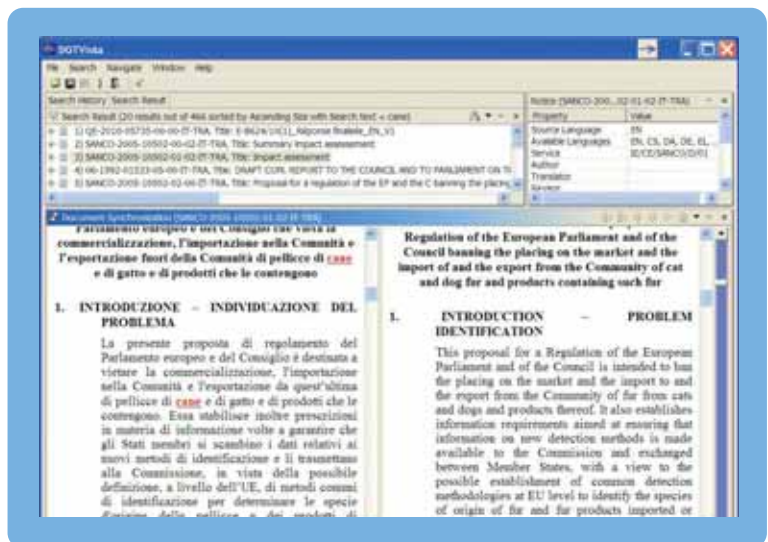


DGT Vista

DGT Vista is a document search and viewing engine. It contains all incoming (mainly original texts) and outgoing documents (mainly translations) from and to every Commission department since 1994. The DGT Vista interface offers translators a range of search criteria (document number, author, requesting service, title or even contents of the text), and enables them to find virtually any document within a matter of seconds.

Each document has a kind of identity card containing all key information.

DGT Vista's particular features (short response time, bilingual parallel scrolling, document downloading from the database into the word-processing system, full-text search facility) have turned it into a very powerful translation aid. The interface also

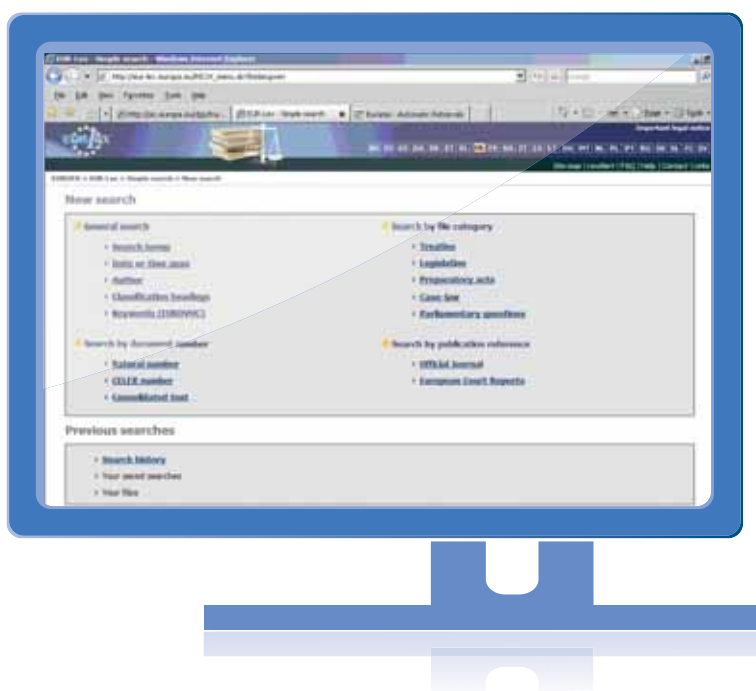


makes it possible to send two documents directly to Euramis for alignment (see the specific chapter on Euramis below).

EUR-Lex

EUR-Lex is an online repository of published EU legislation. It can be accessed by anyone free of charge (at <http://eur-lex.europa.eu>). It contains the treaties, secondary legislation and preparatory acts in all official EU languages, as well as national implementing measures and case-law of the Court of Justice of the European Union. It is also possible to consult the *Official Journal of the European Union*.

Our translators can access EUR-Lex via the web for consultation purposes, as well as indirectly through the Euramis interface (see the specific chapter below), in order to create a translation memory based on EUR-Lex content. Further ways to query EUR-Lex are via Quest II and DocFinder (see explanations further below).



Translation tools

DG Translation has four main types of translation aid: **terminology tools**, **translation memory technology**, **machine translation** and **speech recognition**. There are two levels of translation memory — central and local — each using a different system.

Terminology tools

IATE

Since 2005, IATE (InterActive Terminology for Europe) has been the terminology database shared by all EU institutions and bodies. Since 2007, IATE has also been accessible to the public (at <http://iate.europa.eu>).

Search results indicate the institution that created the entry and the context in which the term is used. Entries also have a reliability code, a numerical value of 1 to 4, 4 meaning 'very reliable' and 1 meaning 'reliability not verified'.

The database contains over 8.7 million terms and half a million abbreviations, covering all official EU languages, as well as Latin. IATE is developed and maintained by an interinstitutional team, but the contents are managed by individual language departments themselves. Every translator in DG Translation can create entries in any language in the database. Mother-tongue terminologists then validate the entries to ensure that the contents of the database are of high quality.

Specific source-language terms or abbreviations and their equivalents in other official EU languages can be searched for in IATE. Searches can also be refined by specifying the domain or context in which a term is used.



DocFinder

During the translation process it is often necessary to consult reference documents for background information. Or, if a related document has been translated already into a specific target language,

the translation might prove useful for some terminology questions.

DocFinder helps to quickly locate and display a document based on references

in the text being translated, such as 'Regulation (EC) No 44/2006'. With DocFinder the full text of a reference document can be retrieved within seconds; highlighting the document reference in the Word document, and

then clicking the DocFinder Word macro button is all that is needed. Alternatively, the document reference can be entered manually in the DocFinder web interface. DocFinder is also used by other EU institutions.

Multidoc 2.0

Multidoc holds a well-organised collection of links to external sources of information, which is added to and maintained collectively by the translators. While translating a document, translators often find good information sources (references, glossaries) on the Internet. In order to save time when looking for the

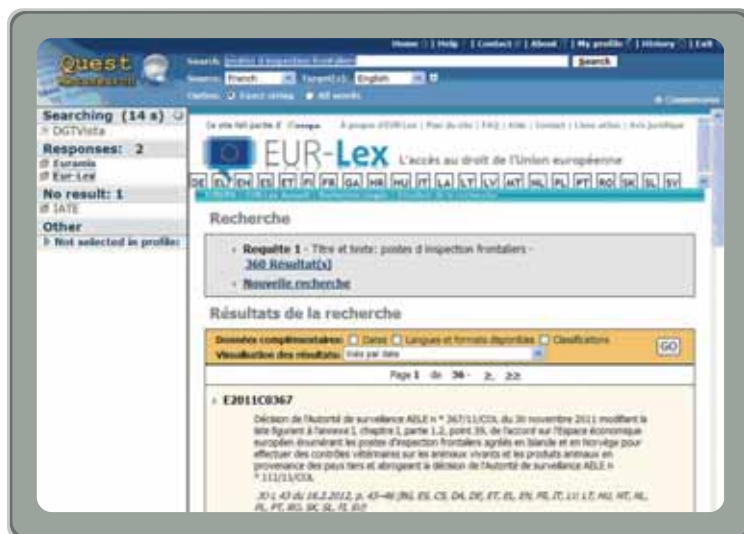
same information source a second time, or in order to share the link with other translators, Multidoc allows bookmarks to that source to be added. Accessing those bookmarks is very easy, and publishing the information and sharing with all colleagues is immediate.

Quest II

Quest II is a metasearch tool designed to drastically reduce the time it takes translators to find solutions to terminology problems. Quest II enables translators to search a multitude of DG Translation's internal and public terminology sources in the time it would normally take to search a single source.

The web interface was developed in DG Translation with a view to centralising,

simplifying and speeding up terminology searches. Our translators and those of other institutions can select the source language and up to three target languages on screen, and specify exactly which databases they wish to include in the search. A Quest button in the Word toolbar makes it possible to launch searches for selected terms directly from a Word document. Like DocFinder, Quest II is used across institutions.



Translation memory technology

The Euramis central translation memory

DG Translation has developed a huge central translation memory as part of the Euramis project (see specific chapter below), with the underlying idea of providing facilities for genuine data sharing between all our staff.

The Euramis central translation memory is not used directly during the translation process; it is merely a database layer which is accessed to retrieve or store data processed locally by the translators, using a CAT tool in stand-alone mode.

At present, the Euramis central memory contains around 645 million segments covering all official EU languages.

English and French are the most commonly used source languages, reflecting the fact that nearly all documents sent to us for translation are written in one of them. When it comes to target languages, retrievals are more evenly distributed. Automated Euramis pre-processing is carried out on original documents in various formats.

The local memory

Our translators use a CAT tool to work locally with the content of Euramis' translation memories. This CAT tool was selected following an interinstitutional call for tenders and was customised to meet the specific needs of all European institutions' translators.

The tool gives translators access to all language and phraseology resources from a local translation memory. When the user enters an original text, similar or identical segments from previously translated texts pop up as translation suggestions for the job in hand.

DG Translation has defined a given set of metadata (translator, document number, year and client) so that every segment gets a specific label in the translation memory.

The CAT tool is particularly useful since a high proportion of legislative and preparatory documents are based on previous texts or existing legislation. It is mainly used as a front-end for the local and interactive processing of data which is retrieved from, or is to be saved in, the Euramis central translation memory.

Machine translation

The principle of machine translation (MT) is well known: a document is roughly translated from a source language into a target language on the basis of a system of dictionaries and linguistic rules or by using statistical techniques.

Machine translation as offered by DG Translation is used mainly for the three following purposes by the EU institutions' administrators and translators.

- **Browsing.** Capable of translating a large number of pages per hour without human intervention, machine translation gives rapid access to information in languages which requesters do not know.
- **Drafting** in a language other than the requester's mother tongue or main language. Some administrators prefer to write a text in their own language first, request a machine translation and then correct the output.
- **Translating** is the principal reason for requesting machine translation by our translators. For translation purposes, the raw machine output is always edited.

The rule-based ECMT system (European Commission Machine Translation) started development in 1976 and was accessible to both translators and administrators until the end of 2010. It was used to translate up to about two million pages per year. The coverage of ECMT was limited to 10 languages (28 language pairs) and further development was abandoned due to the difficulty of extending the coverage to the many more language pairs needed today.

Since 2010, DG Translation has been responsible for the MT@EC project that aims to develop a common machine translation service to be offered by the European Commission. The services being developed will be used by European and national public administrations and will be customised to their specific needs. This new system will make use of data-driven (statistical) approaches, but depending on the needs, rule-based, hybrid and commercial components can also be integrated. It will offer better quality of output, i.e. better translation. It will also offer better quality of service, for example many more languages in the initial system and the possibility to develop new language pairs and customised solutions to fit the specific needs of users in a flexible and cost-efficient way.

As of mid-2011, translation engines for more than 50 language pairs have been built and are undergoing intensive tests by our translators. On busy days, more than 10 000 pages are sent through the system, and depending on the length of the documents submitted, the requested translations will typically be available within only a few minutes. The raw output is given to translators for post-editing to bring it up to the usual high quality. The correction of the raw translations also helps to improve the quality of the MT engines. After further improvements, the service will be available for daily usage throughout the Commission and subsequently opened up to other potential users.

Speech recognition

A number of our translators use speech recognition software. This allows users to dictate text directly onto their computer in a natural, continuous way, achieving a high degree of accuracy and efficiency. The software is a real time-saver for translators, because they no longer need to type a large part of their work, or have it typed (thus saving secretaries' time as well). The ergonomic and health benefits are also obvious, as adverse physical

effects associated with intensive typing and mouse use are reduced.

Uptake of speech recognition technology in DG Translation has been limited by the fact that this technology has been developed by vendors for only some of the EU official languages. Tests are being carried out with other products to try to find suitable software for other languages as well, but the supply is limited in this field.

Euramis

Euramis concept

Euramis (European Advanced Multilingual Information System) refers to a series of client-server applications which provide access to a variety of services in the field of natural language processing. Euramis is based on the following principles:

- central storage of linguistic resources with a view to data sharing (translation memories);
- mass processing of linguistic data;
- integration of various language applications and services with a view to giving one-stop access to, for example, translation memories;
- workflow automation.

The Euramis project was launched in 1995. The underlying idea was to relieve translators of the more repetitive work and to achieve greater consistency in language and methodology, thus contributing to better quality assurance.

Euramis is unique because:

- there is no other translation tool in the world capable of dealing with such a huge volume of material (translation segments);
- it is truly multilingual;
- it is incredibly fast; and
- it is the backbone of all the institutions' translation tools.

Quality assurance is a major concern for all the EU institutions' translators. Therefore, to further improve consistency and allow genuine data sharing between different institutions, Euramis can be accessed by translators from the Parliament, the Council, the Court of Justice, the Court of Auditors, the European Economic and Social Committee, the Committee of the Regions and the Translation Centre for the Bodies of the European Union.

How does Euramis work?

Euramis services are launched automatically as part of the document workflow, so in most cases the data needed by translators is available when they begin work. However, requests can also be submitted manually through the Euramis web portal. This enables the user to submit instructions and any files

to be processed. Euramis then fulfils the request by retrieving the data requested from the sources shown in the illustration below and returns the results to the user by e-mail or places them on a dedicated server.

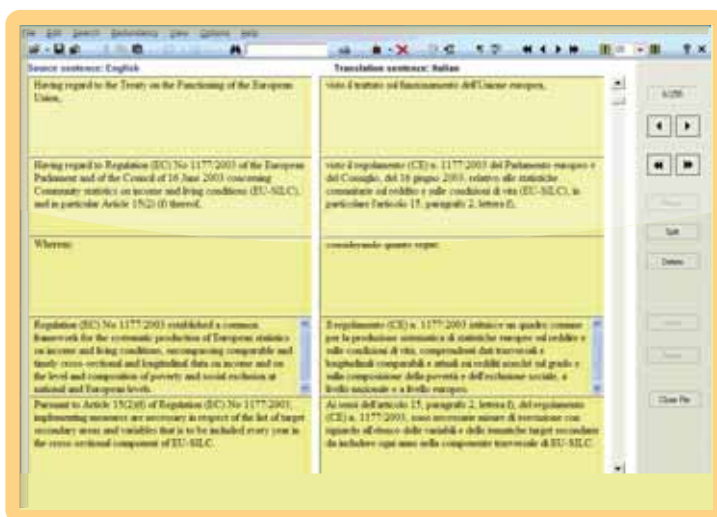
Alignment Editor

The Euramis Alignment Editor is simple to use and offers all the functions users might need when correcting an alignment.

Sentence cells can be deleted, merged or split. It is normally assumed that the sequence of source and target sentences is parallel, but the cut, copy and paste functions can also be used. It is also possible to spell-check the target text.

Euramis Alignment Editor also makes it possible to automatically check for redundant entries and change or add attribute values (metadata) to a given document.

Finally, corrected alignments can be saved directly to a Euramis database.



Central memory retrieval

A retrieval from the Euramis central translation memory is launched on the basis of an original document.

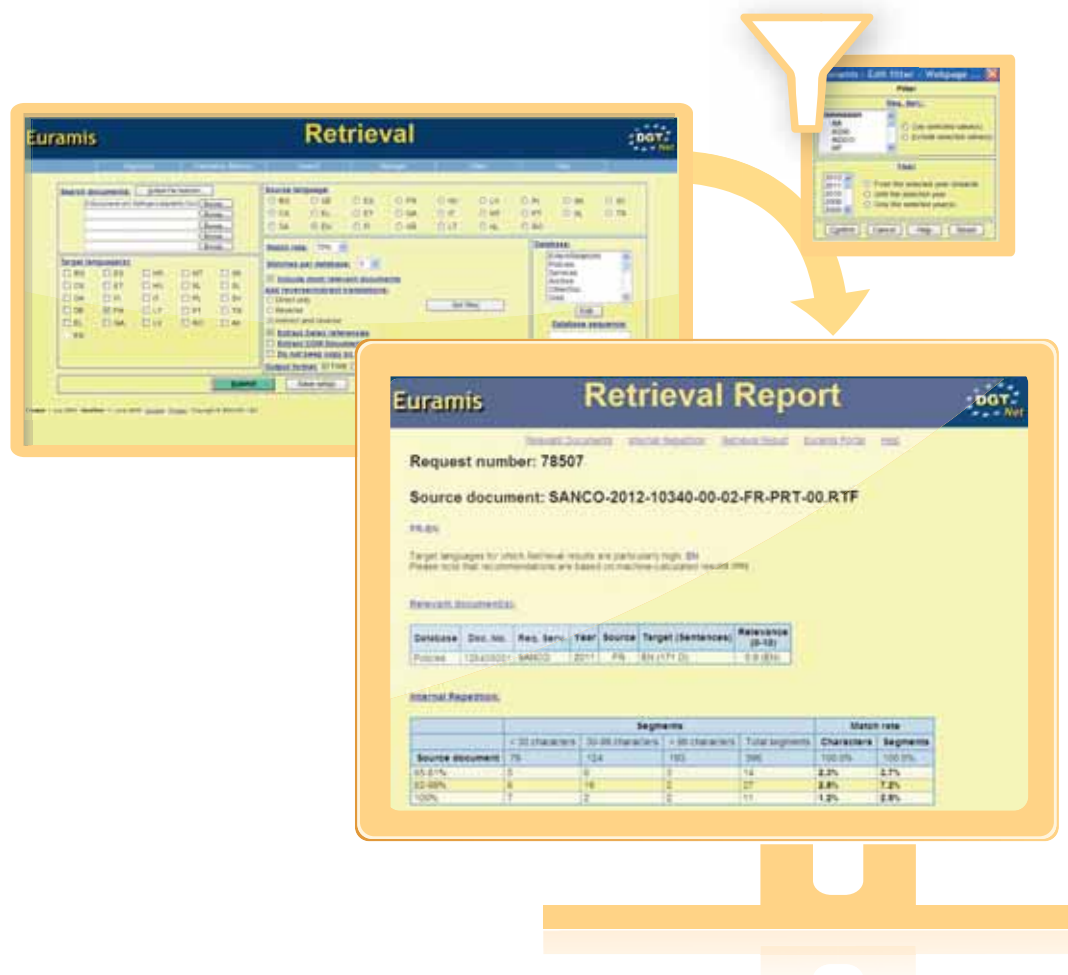
There are several search and selection criteria and other options.

The user can opt for two different output formats.

- **TMX file:** linguistic content to be imported into the local CAT tool for interactive translation. This format allows for inclusion of most relevant documents (when a certain number of translation units come from the same

document, the whole document is retrieved and added to the TMX file).

- **Word output:** a pre-translated file. Sentences are automatically replaced in the Word document with the best matches found. Colours are used to indicate whether they are a perfect match or fuzzy match, or the original text has been left. Comments contain the metadata. It may be used for freelance translators and can serve as a visual aid for decision-makers (how much text is actually new and which tools would be best to use for a given document).



Translation memory retrieval can also be refined using a given set of filters corresponding to the metadata used during the translation process.

For each request, a report is created containing detailed information about the results retrieved.

Combined services

Euramis offers several combined services, for example 'Celex + alignment' and 'DGT Vista + alignment'. The underlying

idea was to integrate various language services with each other and provide for one-stop access.

Save

Euramis allows for saves to the central translation memory (TMX files produced

by a corrected alignment or by exporting a local translation memory).

Document Search

This feature lets users search for specific documents in Euramis translation memories. Once a document has been found, users can view it, download it, or send feedback to the database manager.

Sentences with the same attributes (document number, requesting service, document type, observation, year and language) are regarded as pertaining to the same document.

Euramis online concordance

Euramis Concordance searches Euramis translation memories for the text entered in the text box. If one or more matches are found, a result table is created which contains the sentences found on the left-hand side and their translations on the right-hand side.

- view all sentences from the document (source and target);
- retrieve all sentences from the document;
- send feedback on the document to the database manager.

Sentences from the same document (i.e. with the same attributes) found in the same translation memory are grouped together under a heading identifying the document.

There is also an ‘advanced search’ function which offers more scope for fine-tuning searches, but takes considerably longer.

Each document heading contains 'Show.doc', 'Download' and 'Feedback' buttons which users can click respectively in order to:

Euramis Concordance is a very useful and heavily used tool, as proved by the over 50 000 queries processed every day.



What's next?

DG Translation will continue to adapt its tools and workflow to new developments in translation technology, focussing mainly on:

- further workflow automation;
- further integration of language applications and services;
- further integration at interinstitutional level;
- the creation of a new desktop environment, Mandesk, which will make the workflow more flexible, integrated and logical.



You find more information and publications on the Directorate-General for Translation of the European Commission and our contact information on our website
<http://ec.europa.eu/dgs/translation>

*Europe Direct is a service to help you find answers
to your questions about the European Union.*

Freephone number (*):

00 800 6 7 8 9 10 11

(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

More information on the European Union is available on the Internet (<http://europa.eu>).

Luxembourg: Publications Office of the European Union, 2012

ISBN 978-92-79-23310-4

doi:10.2782/50856

© European Union, 2012

Reproduction is authorised provided the source is acknowledged.

Printed in Belgium

PRINTED ON WHITE-CHLORINE-FREE PAPER



Publications Office

ISBN 978-92-79-23310-4



9 789279 233104