Translation tools and workflow

European Commission

Foreword

Karl-Johan Lönnroth
Director-General
Directorate-General
for Translation
European Commission

Nowadays, translating is not just a matter of finding the right terms and expressions, but also of using the right technology. The Directorate-General for Translation of the European Commission, which is the largest and most complex translation service in the world, has therefore incorporated a unique set of translation tools into its complex workflow.

Thanks, among other things, to translation memory technology, the Directorate-General for Translation has genuine data sharing: translators can avoid retranslating what has already been done and give other colleagues the benefit of their work.

With all the requisite data at hand, translators can concentrate on their 'core business', searching for the right terms and expressions in a more efficient and less time-consuming way and practising the art of translation without the drudgery of having to perform repetitive tasks.

This booklet sets out to show how, under ideal conditions, documents are processed by the Directorate-General for Translation, from the day they arrive until the day they are finalised and delivered to the requesting department.

Manuscript completed in January 2009

European Commission
Directorate-General for Translation
Communication and Information Unit
B-1049 Brussels

A great deal of additional information on the European Union is available on the Internet. It can be accessed through the Europa server (http://europa.eu).

Luxembourg: Office for Official Publications of the European Communities, 2009

ISBN 978-92-79-07952-8

doi: 10.2782/2642

Printed in Belgium

Printed on white chlorine-free paper

Of course, what is presented here is the model scenario, but this is not so far removed from today's reality on the translation market: all translators are liable to be affected by the new developments sooner or later. Indeed, there are various tools on the market which offer some of the functions discussed in these pages, and a number of projects are being conducted along lines similar to our own.

Even at the individual level then, document management and workflow are bound to become increasingly important.

We hope that specialists interested in the computerisation of translation support functions and document workflow for professional purposes will find ideas in this booklet which they can consider adapting to their own situations. The general reader, too, will gain an idea of how the Directorate-General for Translation uses computers to deliver finished language products to the European Commission as quickly, accurately and cost-effectively as possible.

1 The Translation DG of the European Commission

THE TRANSLATION DG OF THE EUROPEAN COMMISSION

For a better understanding of how we work, let us first look briefly at how the Directorate-General for Translation (or DGT) is organised and at the constraints it has to work under ([1]). With approximately 1 750 staff directly involved in translation and 600 support staff, DGT is located in both Brussels and Luxembourg, and is the largest translation service in the world. Other Community institutions and bodies (Council, Parliament, Court of Justice, European Economic and Social Committee, Court of Auditors, etc.) have their own translation departments, whereas the various specialised decentralised Community agencies send their translation work to the Translation Centre for the Bodies of the European Union.

DGT ([2]) is arranged in a **linguistic structure:** each official EU language has its own language department which is organised in 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 (apart from some specific exceptions). There is also a central demand management unit serving all language departments.

([1]) For more information about the organisation and activities of DGT see also 'Translating for a multilingual community' (http://ec.europa.eu/dgs/translation/bookshelf/brochure_en.pdf).

([2]) See also http://ec.europa.eu/dgs/ translation/index_en.htm

DGT is located in both Brussels and Luxembourg, and is the largest translation service in the world. Its total production is some 1.7 million pages a year.

The translator's needs

2

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

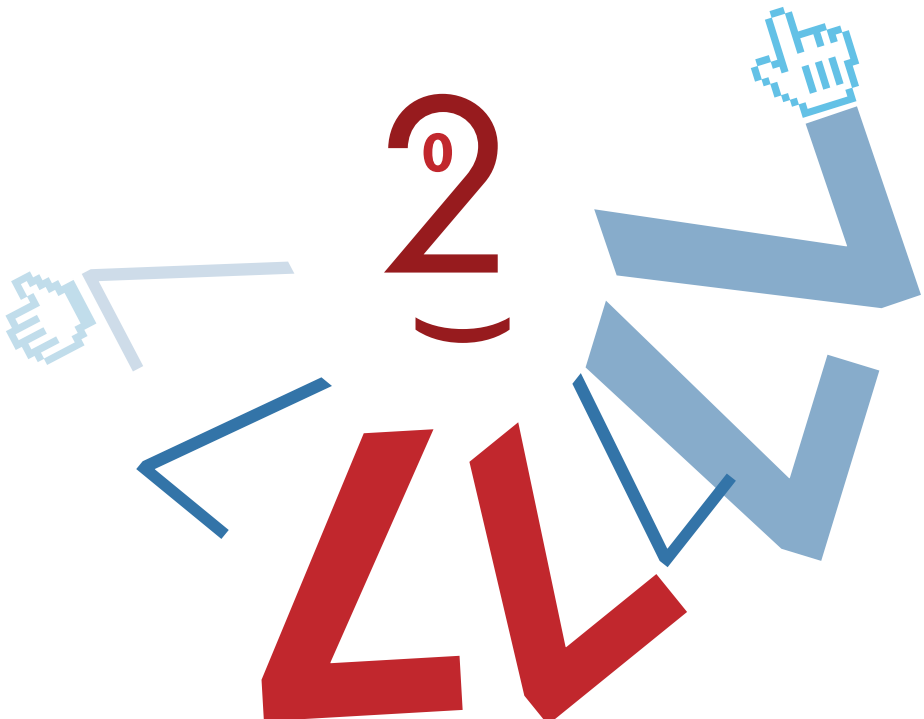
Irrespective of their preferred working methods, all translators' needs are basically the same:

- **appropriate terminology** (dictionaries, glossaries, terminology databases, etc.);
- **reference documents** (paper, electronic archives, aligned texts, etc.);
- **capability to reuse previously translated texts** (translation memories, electronic archives, etc.); and
- **central and local assistance.** The role of secretaries has evolved from typist into 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. At the central level, assistance is provided by the helpdesk, the alignment and pre-processing team, and at the local level, within the language units.

3 DGT language resources

To perform its tasks, DGT has a wide variety of language resources at its disposal:

- **terminology** in many different forms (multilingual libraries, terminology databases, etc.); at desktop level, terminology searches are mainly performed via IATE (the interinstitutional database) and Quest (one-stop access to a series of general-interest terminology databases);
- **translation memories** held centrally by Euramis (see Section 5), thus enabling genuine data sharing;
- **texts** as such to be retrieved from DGT's internal archiving system (called DGTVista) or from any other source;
- **machine translation,** which, at the European Commission, is used not only as a browsing tool but also as a genuine translation aid, and can thus be regarded as a fully fledged language resource.



4 DGT tools

Administration and documentation tools

Poetry

Poetry is the software used for the electronic transmission of translation requests from clients to DGT. The web interface constructs an electronic folder containing the translation request, the original document for translation, and any reference documents required, all of which can then be sent to DGT in one go.

The advantages are numerous and obvious:

- faster transmission,
- integration into DGT's electronic archiving system,
- availability of original and reference documents in electronic form,
- improved electronic workflow, etc.

Suivi

Suivi is the software used for the electronic management of translation requests within DGT. The program monitors the progress of a document and sends translations back to requesters.

Dossier Manager

Dossier Manager is the interface for translation management.

When the identification number corresponding to the translation request is entered in Dossier Manager, the interface gives access to the original document and to all files needed during the translation process, including:

- reference documents,
- pre-processing files,
- document comparisons,
- ongoing translations, and
- documents already released.

This is the tool for creating a translation, and it includes an alert and note function for communication between translators working on the same translation project. Dossier Manager is also DGT's electronic archiving system.

DGTVista

DGTVista, a document search and view engine, contains all incoming (mainly original) and outgoing (mainly translation) documents from and to every directorate-general and service in the Commission since 1994. Its web interface offers users a range of search criteria (document number, author, requesting service, title or even contents of the text), thus enabling them to find virtually any document within a matter of seconds.

Each document has something called a notice, which is a kind of identity card containing all key information.

DGTVista'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 new web interface also makes it possible to send two documents directly to Euramis for alignment.

EUR-Lex

EUR-Lex ([3]) — formerly Celex — is an online repository of published EU legislation. It can be accessed by the general public free of charge, allowing them to consult the *Official Journal of the European Union*. 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 Communities.

DGT translators can access EUR-Lex via the web for consultation purposes, as well as indirectly through the Euramis interface, in order to create a translation memory based on EUR-Lex content.

([3]) **http://eur-lex.europa.eu**

Translation tools

DGT has three main types of translation aids: terminology tools, translation memory technology and machine translation. There are two levels of translation memories: central and local. Both use different systems.

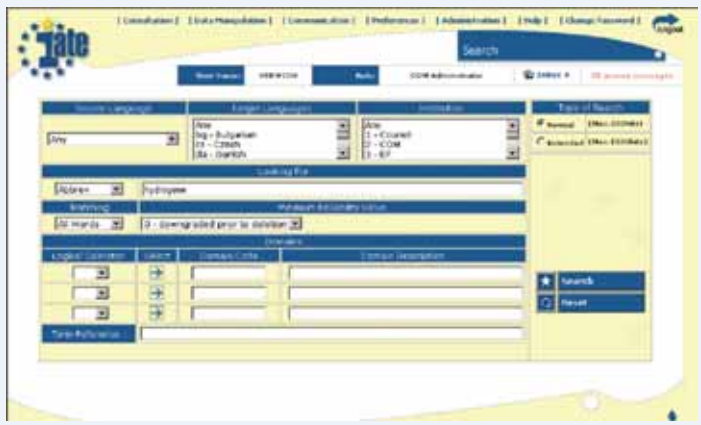
Terminology tools

IATE

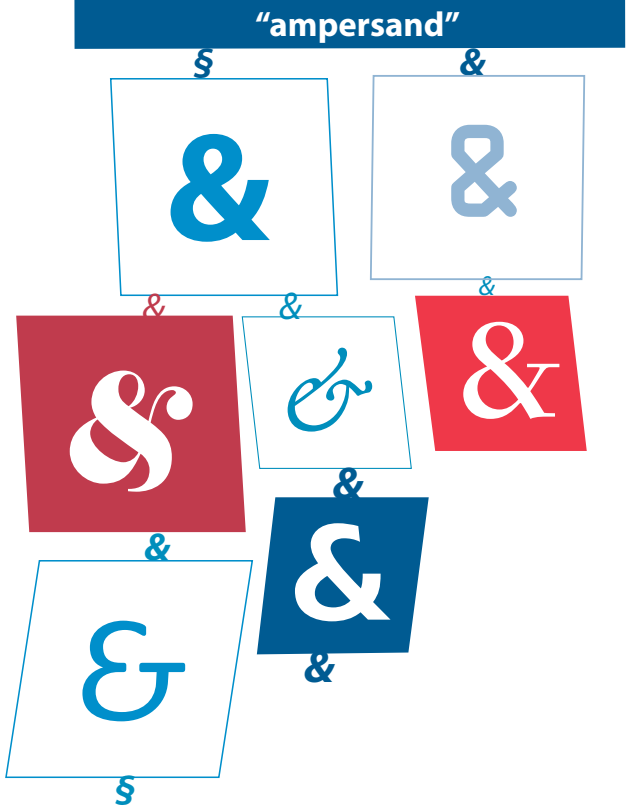
IATE (InterActive Terminology for Europe) is the shared terminology database of all EU institutions and bodies. It became fully operational within the European Commission at the beginning of 2005. It contains over 8 million terms and 560 000 abbreviations, and covers all official EU languages, as well as Latin. The database is developed and maintained by an interinstitutional team, but the contents are managed by the language departments themselves. Every translator in DGT has the right to insert entries into the database. Mother-tongue terminologists validate all entries to ensure that the contents of the database are of high quality.

IATE can be searched for a specific source language term or abbreviation and its equivalent in any one of the 23 other languages, or indeed in all of them. Searches can also be refined by specifying the domain or context in which a term is used.

Search results indicate the institution that created the entry and the context in which the term was used. Entries also have a reliability code, where four stars signify 'very reliable' and one star means 'reliability not verified'.

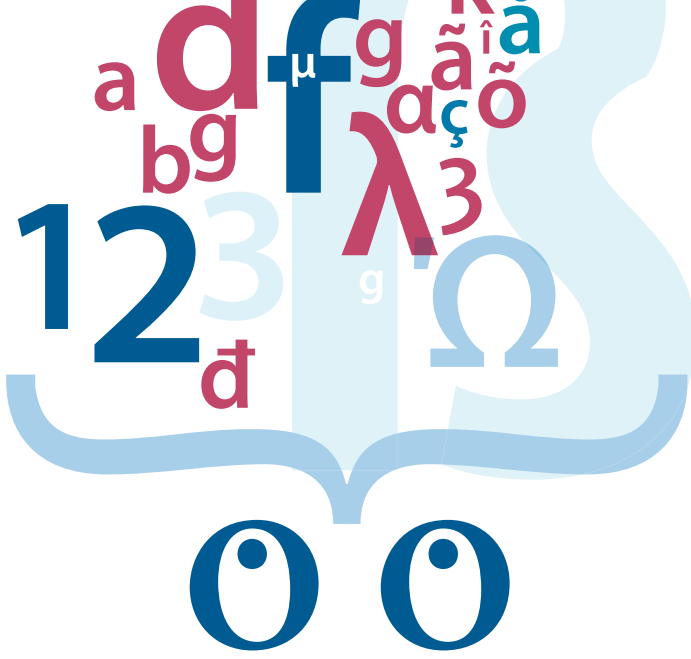


In June 2007, IATE was opened to the public. It can be consulted online at http://iate.europa.eu



Quest

Quest is a metasearch tool designed to drastically reduce the time it takes translators to find solutions to terminology problems. Quest enables translators to search about 30 DGT internal and public terminology sources in the time it would normally take to search a single source. The web interface was developed in DGT with a view to centralising, simplifying and speeding up terminology searches. Translators from DGT and other institutions can select on screen the source and up to three target languages, and specify exactly which databases they wish to include in the search. A Quest button on the toolbar makes it possible to launch searches for selected terms directly from Word.



Translation memory technology

The Euramis central translation memory

A huge central translation memory was developed as part of the Euramis project (see Section 5), the underlying idea being to provide facilities for genuine data sharing between all DGT 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 Translator's Workbench and/or Word or TagEditor as front ends.

At present, the Euramis central memory contains more than 205 million phrases ('segments') in all official EU languages.

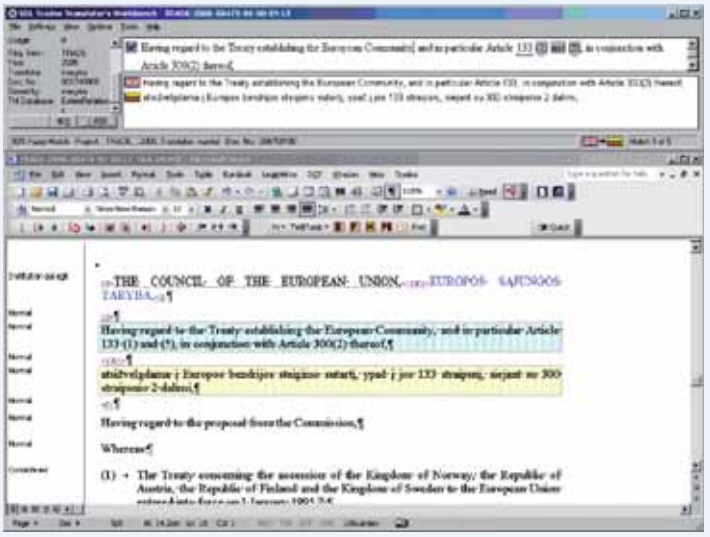
As source languages, English, French and, to a lesser extent, German are the most commonly used, reflecting the fact that nearly all documents produced by the Commission are written in one of these languages. When it comes to target languages, retrievals are more evenly distributed. Automated Euramis pre-processing (retrieval) is carried out on all original documents in Word, HTML and XML format.

SDL Trados Translator's Workbench — the local memory

Translators at DGT use an integrated translation support tool to work with the content of Euramis' translation memories. SDL Trados Translator's Workbench (TWB) was selected following an interinstitutional call for tenders and was then customised to meet the European institutions' specific needs.

It was chosen mainly for its coverage of all EU languages and its capability to handle TMX files (Translation Memory eXchange), the standard exchange format for translation memories used by Euramis.

TWB 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.



In TWB, DGT has defined (via project settings) a given set of attributes (translator, document number, year and client) so that every segment gets a specific label in the translation memory.

TWB 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

ECMT

The principle of machine translation (MT) is well known: a raw translation of a document, from a source language into a target language, is made on the basis of a system of dictionaries and linguistic programs.

ECMT (short for European Commission Machine Translation) started development in 1976 and is accessible to both translators and administrators. It has three main uses, to aid the tasks listed below.

- **Browsing:** Capable of translating up to 2 000 pages per hour, it 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 officials prefer to write a text in their own language first, request a machine translation and then correct the output.
- **Translating:** This is the principal reason for requesting machine translation within DGT. When used as the basis for a fully fledged translation of a document, the raw machine output must be edited. The amount of correcting required varies according to the text type (letter, minutes, manual, etc.).

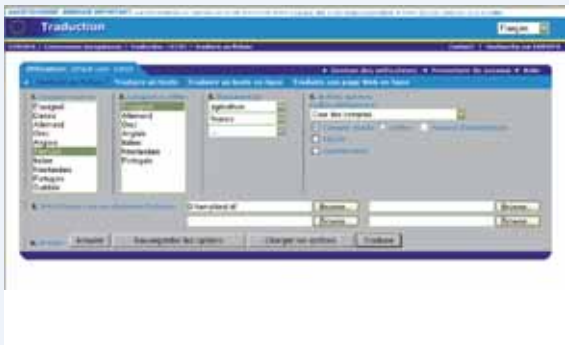
Total MT production was 1 963 991 pages in 2008. Amongst all institutions, the Commission was the main user with 1 789 770 pages. DGT accounted for 1 513 825 of these pages, due to the fact that documents are systematically pre-processed with machine translation for certain language combinations. The actual use of MT as a translation aid in DGT is difficult to calculate, but it lies somewhere between 23 and 50 %. MT is thus used in the Commission both as a genuine translation aid and as an administrative support tool.

Currently, ECMT offers translation for 18 operational language pairs (or combinations):

From English into	From French into	From Spanish into
Dutch	Dutch	English
French	English	French
German	German	**From German into**
Greek	Italian	English
Italian	Portuguese	French
Portuguese	Spanish	**From Greek into**
Spanish		French

The quality of a given language pair mainly depends on the length of development and language similarities. However, MT quality is also heavily influenced by the original document: if the source text is unclear, contains typing errors or is syntactically complex, the result will definitely be poor. ECMT's main strengths are the coverage of its dictionaries — which have been adapted to cover the Commission's areas of work — and, of course, speed.

MT can be requested via the user's e-mail application or a user-friendly interface on IntraComm, the Commission's Intranet. It can also be accessed via Euramis (see Section 5).



Voice recognition

Some 300 translators in DGT are currently using some kind of speech recognition tool. This software 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 by DGT has been limited by the fact that this technology has been developed by vendors for only 9 of the current 23 EU official languages. Tests are being carried out with other products to try to find a suitable software for other languages as well, but the supply is limited in this field.

5 Euramis project

The Euramis project was launched in 1995 following a call for tenders for the 'Development of multilingual tools and their integration into multilingual services'. 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 segments**;
- unlike most translation memory systems it is **truly multilingual**;
- it is **incredibly fast**; and
- it is the **backbone** of all the institutions' translation tools.

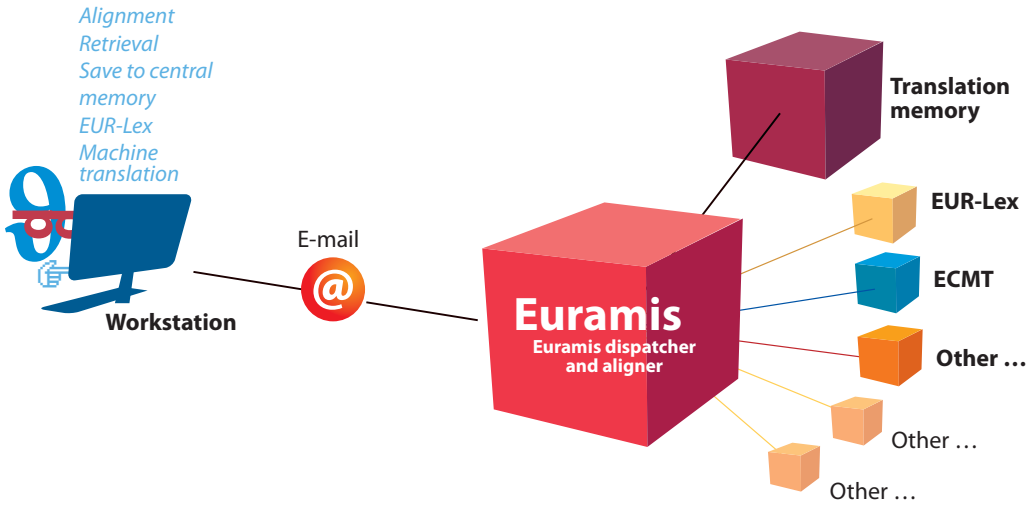
Quality assurance is a major concern not only for DGT but also for the other institutions. Therefore, to further improve consistency and allow genuine data sharing between translators working for different institutions, Euramis can now be accessed by users in the Council, the Court of Auditors, the Court of Justice, the Committee of the Regions, the European Economic and Social Committee, the Parliament and the Translation Centre for the Bodies of the European Union.

Euramis concept

Euramis (European advanced multilingual information system) refers to a series of e-mail-based, 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, and machine translation;
- workflow **automation**.



How does Euramis work?

Euramis services are launched by means of the Euramis web interface.

At workstation level, the user creates a request (instructions and files to be treated) using a web interface. A service dispatcher reads the command files and, after going through the processes shown in the illustration, the final results are returned to the user's mailbox.

Machine translation

Euramis gives direct access to machine translation. A machine translation request can also be combined with a retrieval from the Euramis central memory.

EUR-Lex download

By typing EUR-Lex document references into the appropriate box of the Euramis web interface, or by attaching a file which contains EUR-Lex references, the user can receive the full titles or text of the corresponding legal acts in one or more EU languages (up to 30 documents) by e-mail.

The download function also allows for automatic alignment of downloaded acts at server level. In that case, the user receives a TWB import file (in .tmx format).

A similar service exists for the main categories of internal Commission documents.

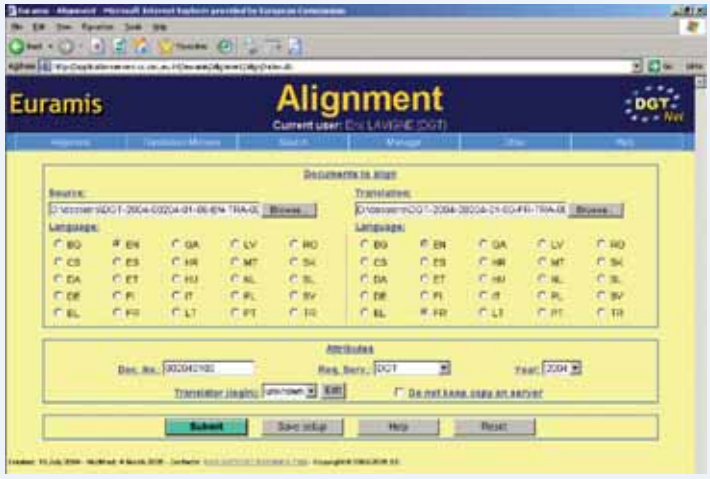
Alignment ([4])

When the Euramis project was launched in 1995, commercially available alignment programs did not offer acceptable quality, hence the decision to develop a highly customised Euramis aligner.

Like most Euramis applications, requests for alignments are launched via the Euramis web interface and the results are received by e-mail.

The user has to define values for a given set of attributes, which are the same as those used in TWB project settings and thus enable individual labelling of segments in the central translation memory.

([4]) An operation consisting of splitting two existing texts in different languages (source and target) into segments, usually sentences, and then placing these segments in parallel. The result of the operation is a single file containing the various segment pairs which, after checking (see Alignment Editor), is fed into the (local and/or central) translation memory.



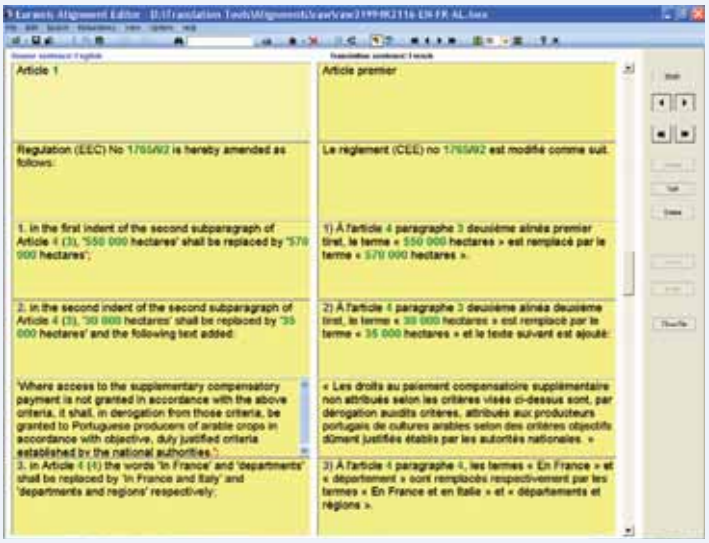
An alignment can be made for:

- instant use during the interactive translation process;
- storage in the Euramis central translation memory for later use.

In both cases corrections can be made using the Alignment Editor.

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.

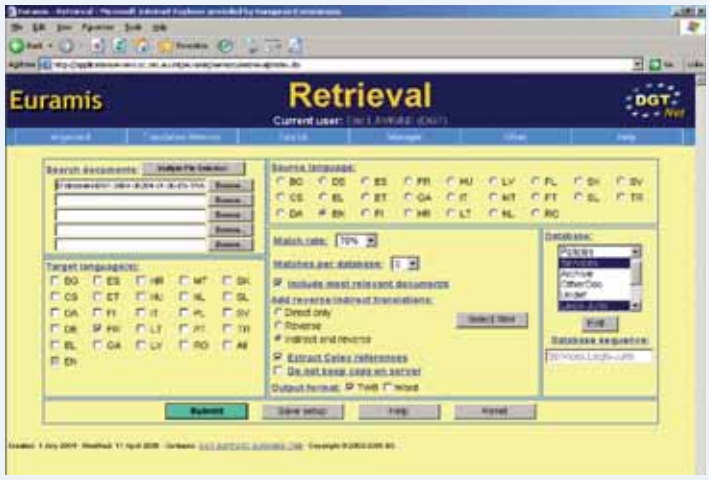
Euramis Alignment Editor also makes it possible to automatically check for redundant entries and change or add attribute values (project settings) to a given document.

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.

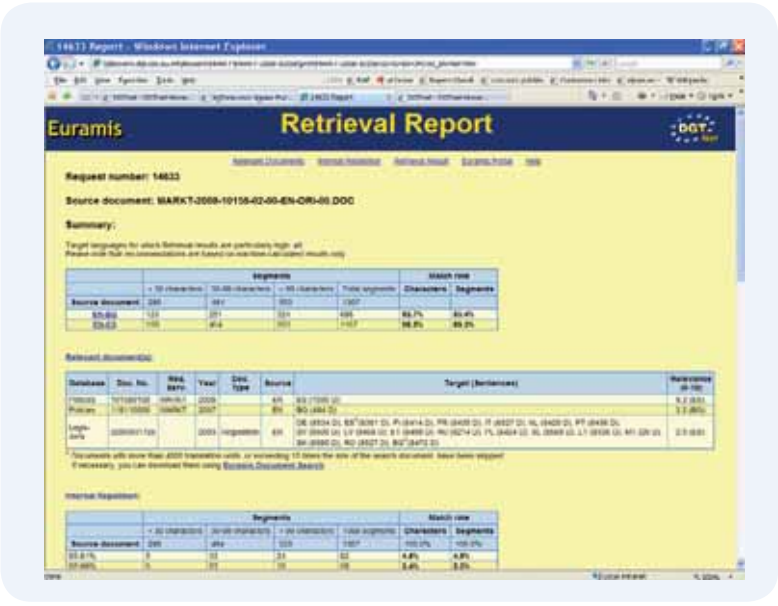


The user can opt for two different formats.

- **TWB import file (.tmx file)**: a kind of 'project memory' to be imported into TWB for interactive translation. This format allows for inclusion of all the most relevant documents (when a certain number of translation units come from the same document, the whole document is retrieved and added to the TWB import file).
- **Word output**: corresponds more or less to the Translate function in TWB. Sentences are automatically replaced in the Word document with the best matches found. Colours are used to indicate status (perfect match, fuzzy match, original text, machine translation) and comments to display project settings. It can be seen as an alternative for translators without TWB 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 project settings used during the translation process.

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



Combined services

Euramis offers several combined services:

- 'Celex + alignment',
- 'DGTVista + alignment',
- 'Retrieval + machine translation',

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

The example below is an illustration of a combined retrieval after TWB import.

In this case, translation memory retrieval has been combined with a machine translation request.

When Euramis does not find any perfect or fuzzy matches in the central memory, it requests a machine translation, which is included at the end of the retrieval result.

In TWB (see illustration below), a machine-translated sentence of this kind is displayed in grey to distinguish it from perfect (green) or fuzzy (yellow) matches.

The advantages compared with retrieval or machine translation alone are obvious: users always get a suggestion while being aware of its status, and they do not have to keep making the same changes in machine translation, since validated proposals from machine translation become perfect or fuzzy matches, respectively, if they reappear later in the text.

Save

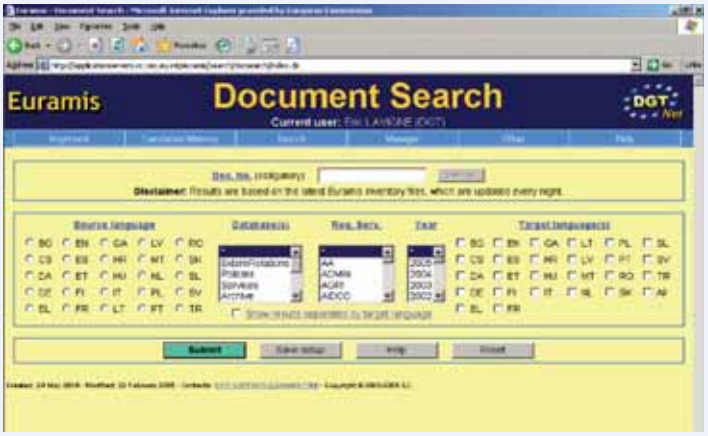
Euramis allows for saves to the central translation memory (aligned files or TWB export files).

Document Search

Document Search lets users search for specific documents in the translation memories of Euramis. Once a document has been found, users can view it, download it, or send feedback to the database manager. Sentences with the same document number, client and year references are regarded as pertaining to the same document.

In order to avoid uncontrolled proliferation of data among the various translation memories, the downloaded sentences are provided in a format which cannot be reimported back into Euramis (unless they have since been reused for translation in TWB).

In order to optimise efficiency, queries are not performed in the actual translation memories, but in the Euramis inventory files, which list the contents of those memories per document, language pair, etc. These inventory files are updated every night. As a consequence, users will not find documents which were imported 'today'.



Euramis online concordance

Euramis Concordance searches the Euramis translation memories for the text entered in the text box. Unlike TWB, it does not have fuzzy matching capabilities, but it is possible to search using wildcards. 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.

Sentences from the same document (i.e. with the same document number, client and year references) found in the same translation memory are grouped together under a heading identifying the document. Each document heading contains 'Show', 'Download' and 'Feedback' buttons which users can click respectively in order to:

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

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

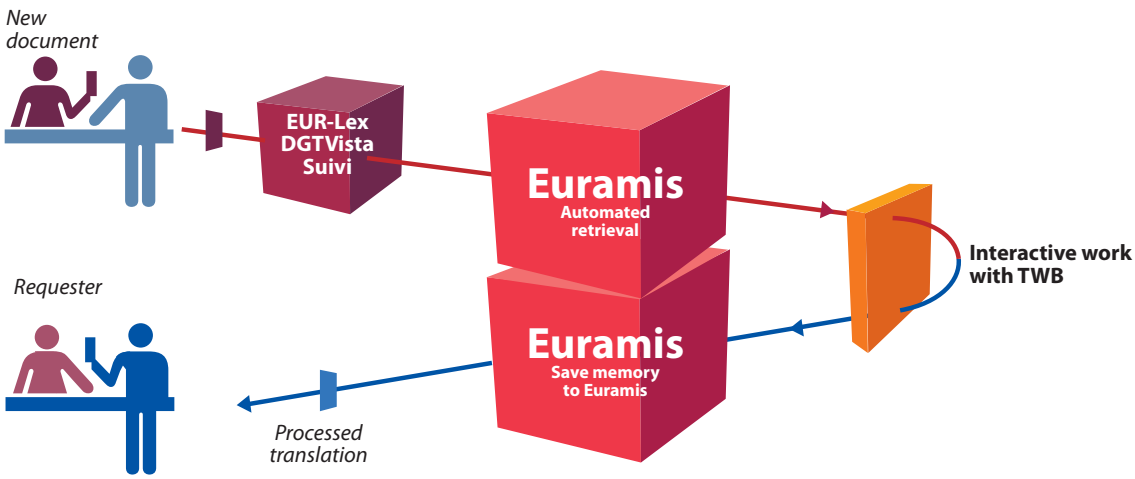
6 Workflow evolution

The current workflow scenario

Unlike in the past, when translation work would arrive directly on the translator's desk, workflow is now rather complex — as can be seen below.

All translation requests are sent by Poetry/Suivi to Euramis for automatic retrieval. Results are automatically stored by Euramis in Dossier Manager. In Dossier Manager, users can access pre-processing files. A macro allows users to automatically create a project translation memory, import selected pre-processing files and fill in the project settings. After interactive translation with TWB, another macro allows users to automatically clean up, export and save to Euramis all translated documents.

Although most of the originals are still in Word, the volume of other formats, such as HTML and XML, has seen a constant increase in recent years, and workflow has been adapted to make it possible to process whole websites sent as a zip file.



7 What next?

As for the future, development efforts will be focused primarily on:

- further workflow automation;
- further integration of language applications and services;
- further integration at interinstitutional level;
- the creation of a new desktop environment, **a Translator's Desktop**, which will make the use of translation technology easier and more logical.

DGT will continue to adapt its tools and workflow to new developments in translation technology. The emergence of new formats where the content is separate from the text formatting process promises to simplify workflow, while relieving the translator of the burden of handling lots of different formats and/or software applications.

8 Annex

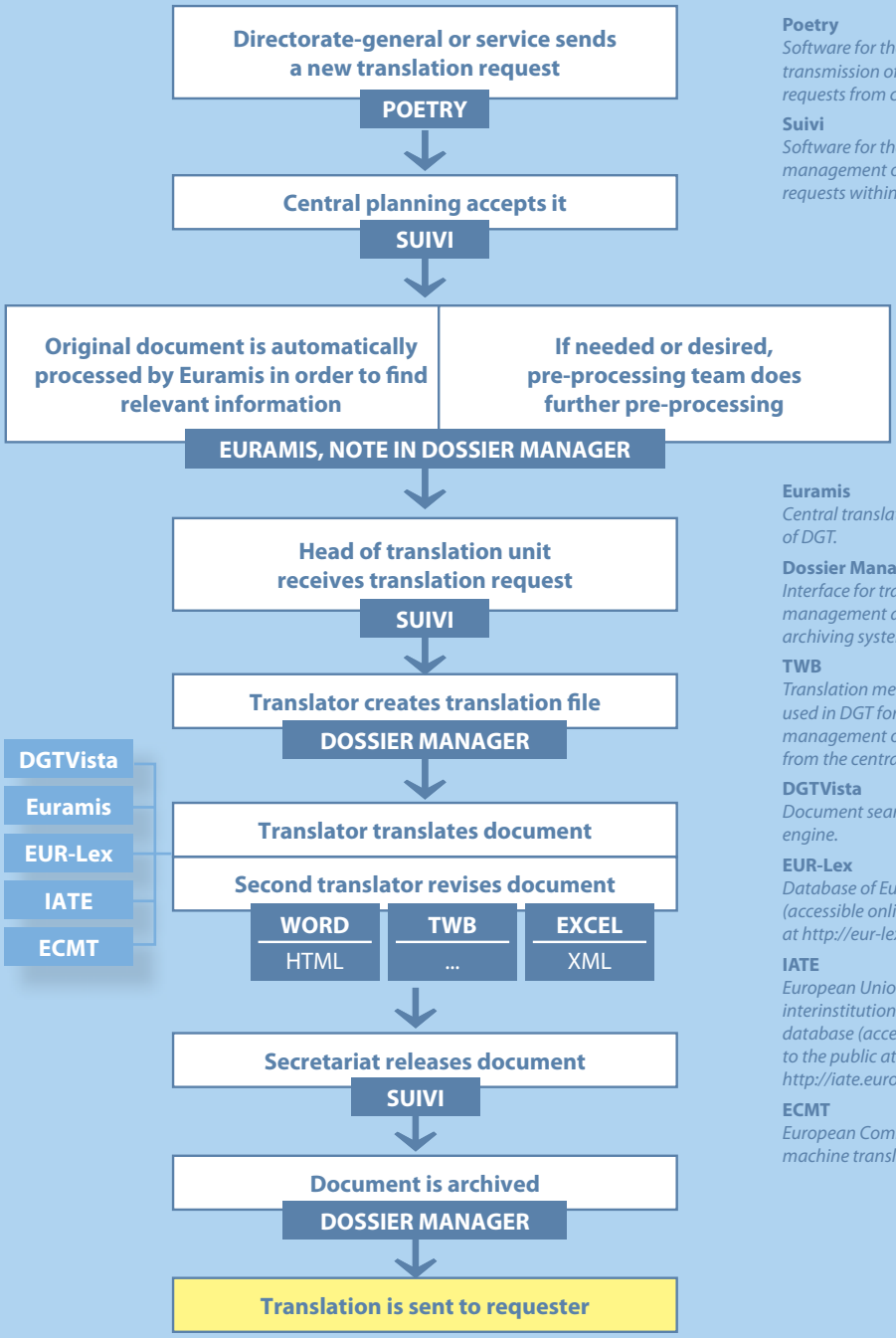
Contacts

For more information on multilingual tools or other matters, please contact:

Multilingual tools
DGT-R-3-SECRETARIAT@ec.europa.eu

Any other matters:
DGT-WEBMASTER@ec.europa.eu

Translation workflow in the European Commission



Poetry
Software for the electronic transmission of translation requests from clients to DGT.

Suivi
Software for the electronic management of translation requests within DGT.

Euramis
Central translation memory of DGT.

Dossier Manager
Interface for translation management and an electronic archiving system.

TWB
Translation memory software used in DGT for the local management of data retrieved from the central memory.

DGTVista
Document search and view engine.

EUR-Lex
Database of European law (accessible online to the public at http://eur-lex.europa.eu).

IATE
European Union interinstitutional terminology database (accessible online to the public at http://iate.europa.eu).

ECMT
European Commission machine translation system.

HC-81-08-467-EN-C

Генерална дирекция „Писмени преводи“
Dirección General de Traducción
Generální ředitelství pro překlady
Generaldirektoratet for Oversættelse
Generaldirektion Übersetzung
Kirjaliku tõlke peadirektoraat
Γενική Διεύθυνση Μετάφρασης
Directorate-General for Translation: http://ec.europa.eu/dgs/translation
Direction Générale de la Traduction
An Ard-Stiúrthóireacht Aistriúcháin
Direzione generale della Traduzione
Tulkošanas ģenerāldirektorāts
Vertimo raštu generalinis direktoratas
Fordítási Főigazgatóság
Direttorat Ġenerali għat-Traduzzjoni
Directoraat-generaal Vertaling
Dyrekcja Generalna ds. Tłumaczeń Pisemnych
Direcção-Geral da Tradução
Direcția Generală Traduceri
Generálne riaditeľstvo pre preklad
Generalni direktorat za prevajanje
Käännöstoimen pääosasto
Generaldirektoratet för översättning

Publications Office
Publications.europa.eu