Denis Dechandon
Head, Language and Technology Support Section
Translation Support Department

TRANSLATION CENTRE FOR THE BODIES OF THE EUROPEAN UNION

DRB Building / Office 2015
12 E, rue Guillaume Kroll – L-1882 Luxembourg
Tel: +352 42 17 11 295 – Fax : +352 42 17 11 290
denis.dechandon@cdt.europa.eu || www.cdt.europa.eu

Brief autobiographical note

Denis Dechandon has over 20 years' experience in translation and linguistics, in office automation (with a programming background) and in various management roles. After getting acquainted with the translation work and its requirements at EU level, he fully committed himself to the definition and implementation of various processes and workflows to provide a better support to linguists and to streamline the working of support teams.

Denis is in charge of a service dedicated to the linguistic and technical support provided to translators, revisers and editors and to the maintenance and enhancement of tools and resources at the Translation Centre for the Bodies of the European Union. Building on human resources, communication, IT tools, computers and management tools, he is committed to further changes and evolutions, such as the co-definition of new services and products and the introduction of machine translation in the translation process in his professional environment.

Viviane Schütz
CAT Tools Coordinator
Language and Technology Support Section
Translation Support Department

TRANSLATION CENTRE FOR THE BODIES OF THE EUROPEAN UNION

DRB Building / Office 2020
12 E, rue Guillaume Kroll – L-1882 Luxembourg
Tel: +352 42 17 11 253 – Fax : +352 42 17 11 290
viviane.schutz@cdt.europa.eu || www.cdt.europa.eu

Brief autobiographical note

Viviane Schütz has been the coordinator for CAT tools at the Translation Centre for the Bodies of the European Union since 2000.

After graduating from Liège University (Belgium) in Germanic languages and literature, Viviane taught German and English, before joining Logos Corporation in the frame of a project financed by the Belgian Ministry of the Walloon Region. She worked on the Logos machine translation system as a linguist in the R&D department. She then worked in customer support at A.L.P. Systems, later at Siemens A.G. in both customer support and development of the METAL machine translation system. Since joining the Translation Centre, she has been managing CAT projects, offering support to internal and freelance translators and has been an active member in interinstitutional committees, CAT tools call for tenders and selection boards.

'Building on processes, workflows, skills and IT tools to speed up the translation process and meet clients' expectations'

This paper reports on the results of an approach put in place at the end of 2010 at the Translation Centre for the Bodies of the European Union.

The Translation Centre for the Bodies of the European Union was created on 28 November 1994 as a rational response to the translation needs of a large number of European agencies and offices. This role was enlarged in 1995, as it was necessary to strengthen interinstitutional cooperation in the field of translation, 'to enable the Centre to carry out, in the medium term, any activities the amalgamation of which has been decided on in accordance with the rules in force'. This enlargement, introduced on 30 October 1995, gave a new dimension to the Centre's key role. Firstly, translation and related activities were extended to those institutions and bodies having a translation service. This type of cooperation is established on a voluntary basis and with the aim of absorbing any surplus work which may arise. Secondly, the amendment to the Founding Regulation gave rise to the Centre's active participation in interinstitutional cooperation with a view to 'rationalising working methods and making overall savings'. The Centre is therefore a full member of the Interinstitutional Committee for Translation and Interpretation (ICTI) and contributes to activities in the following fields: resources, terminology, administration, freelance translation, new working methods and techniques, etc.

In the wake of successive enlargements and as a result of the creation of various EU bodies (agencies, institutes, etc.), and following various technological changes, the growing number of clients, fields, language pairs, software and tools used by authors throughout Europe became a challenge to be addressed for the benefit of clients as well as for all freelance and in-house linguists, as all formats are as a rule accepted and all documents have to be processed according to a fixed grid of deadlines.

In addition to some 430,000 pages of trademarks and designs, over 270,000 pages in 24 official EU languages pass through the hands of some 100 in-house linguists (translators/revisers/editors) and freelance translators (around 1,500 contracts) each year for translation, modification, revision or editing to satisfy the needs of 59 clients whose fields of expertise may be organised into 13 groups:

- Banking,
- Education & culture,
- Energy,
- Environment,
- Foreign & security policy,
- Fundamental rights,
- Health & consumption,
- Networks,
- Property rights,
- Research & innovation,
- Science & technology,
- Social affairs,
- Transport

In view of various difficulties identified and some deficiencies in the preparation of the files sent for translation, a new objective was set, namely the definition and implementation of pre-processing activities, both technical and linguistic, to ease the work of translators and revisers, to ensure a higher quality of deliverables and to guarantee the linguistic processing of all incoming requests overall.

The creation of the Language and Technology Support Section (LTS) in September 2010 was the first major step in meeting the set objective. Five existing specialised teams ('CAT Tools', 'Library', 'Reference material', 'Technical' and 'Terminology') covering five traditional domains were thus brought together in a new entity and had to deal with this new objective together.

Breaking down the barriers between teams helped to create synergies and new collaborations, firstly within the newly created section and then with other services (IT Department, Translation Department, External Relations and Communication Section and Human Resources Section). After an intense exchange period, the first results crystallised while pre-processing our very first project, a complex one of several thousands of pages for one of the EU institutions at the end of 2010.

All teams had suddenly to work together on a common project, collaborate, and exchange views on needs (translation memories, terminology, reference material, formatting and layout issues) and possibilities (tools, resources, approaches, segmentation and organisation of the pre-processing work), thus making the best of individual experiences, skills and knowledge as well as of available tools and resources, in order to deliver the best possible result with a tight deadline.

Nevertheless, major constraints quickly appeared because:

- all teams were (and still are) very small,
- all skills are constantly needed as most of the incoming requests received from clients are not scheduled,
- most of the related skills cannot be easily and quickly transferred,
- both types of pre-processing (i.e. linguistic and technical) require advanced skills in various unrelated areas of expertise.

The conclusion drawn from all this was that there was a need to rationalise and rethink the approach in place before creating LTS to achieve the objective that had been set, even if it meant challenging existing patterns and radically changing attitudes and habits, which made change management absolutely necessary.

The result is that, with originally some 18 staff members (now 16), i.e. linguists, IT specialists, technicians and coordinators, LTS deals with the following questions daily:

- How can in-house and freelance translators', revisers' and editors' work be best prepared?
- How can they be supported best?
- How can the translation process be speeded up without putting quality at risk?
- How can duplication of work be avoided at various stages (e.g. it is more efficient to spend five minutes to fix a technical issue once (i.e. in the source file) than to spend five minutes fixing it in each target file (i.e. 115 minutes in the case of 23 target languages, plus the time needed to clean the related translation memories)?

At the same time, it analyses each incoming service request and considers available useful tools and resources (translation memories, corpus of documents, glossaries) before assigning any documents to the linguists[i]. The same questions led to defining sequences of activities and standardising them as much as possible after having identified all client document types.

Idea after idea, process embryos were defined for systematic pre-processing in line with all products/services proposed, for all clients, document types and formats. LTS thus elaborated a pre-processing process (of source documents for translation/revision/editing), various standard operating procedures ('SOPs') and several work instructions ('WINs') to be used by all LTS members for answering all above-mentioned questions.

All in all, it was a success, and the first pre-processing cornerstone had been laid, which led to the further definition, description and formalisation of pre-processing and the documentation of the related process (see chart below).

```
                    ┌─────────────────────────────────────────────┐
                    │  Linguistic service request or processing   │
                    │                  request                     │
                    └─────────────────────────────────────────────┘

        ┌───────────────────────────────────────────────────────────┐
        │        1 – Check and encode the details of the request     │
        └───────────────────────────────────────────────────────────┘

        ┌───────────────────────────────────────────────────────────┐
        │  2 – Carry out the automated pre-processing activities on   │
        │                   the source document                       │
        └───────────────────────────────────────────────────────────┘

        ┌───────────────────────────────────────────────────────────┐
        │                 3 – Assign tasks to LTS teams               │
        └───────────────────────────────────────────────────────────┘

        ┌───────────────────────────────────────────────────────────┐
        │    4 – Carry out the manual (linguistic and technical)      │
        │          pre-processing activities on the source document   │
        └───────────────────────────────────────────────────────────┘

        ┌───────────────────────────────────────────────────────────┐
        │            5 – Finalise all pre-processing activities        │
        └───────────────────────────────────────────────────────────┘

            ┌─────────────────────────────────────────────────┐
            │ Pre-processed source document and supporting     │
            │ material made available to the Demand Management  │
            │ Section and translators, monolingual reference    │
            │ corpus expanded                                   │
            └─────────────────────────────────────────────────┘
```

It also led to the establishment of the related (6) standard operating procedures:

```
                    ┌─────────────────────────────────────────────┐
                    │  Linguistic service request or processing   │
                    │                  request                     │
                    └─────────────────────────────────────────────┘

            ┌───────────────────────────────────────────────────────┐
            │        1.1 – Acquaint yourself with the request        │
            └───────────────────────────────────────────────────────┘

            ┌───────────────────────────────────────────────────────┐
            │              1.2 – Check the request validity           │
            └───────────────────────────────────────────────────────┘

            ┌───────────────────────────────────────────────────────┐
            │                 1.3 – Record the request                │
            └───────────────────────────────────────────────────────┘

            ┌───────────────────────────────────────────────────────┐
            │  2.1 – Carry out the automated technical pre-processing │
            │          activities on the source document              │
            └───────────────────────────────────────────────────────┘

            ┌───────────────────────────────────────────────────────┐
            │  2.2 – Carry out the automated linguistic pre-processing│
            │          activities on the source document              │
            └───────────────────────────────────────────────────────┘

            ┌───────────────────────────────────────────────────────┐
            │               3 – Assign tasks to LTS teams             │
            └───────────────────────────────────────────────────────┘

┌───────────────────────────────────┐     ┌───────────────────────────────────┐
│ 4a – Carry out the manual technical│     │ 4b – Carry out the manual         │
│ pre-processing activities on the   │     │ linguistic pre-processing         │
│ source document                    │     │ activities on the source document  │
└───────────────────────────────────┘     └───────────────────────────────────┘

            ┌───────────────────────────────────────────────────────┐
            │      5.1 – Make sure that all pre-processing results    │
            │                   have been saved                       │
            └───────────────────────────────────────────────────────┘

            ┌───────────────────────────────────────────────────────┐
            │  5.2 – Add the source document to the monolingual       │
            │                   reference corpus                      │
            └───────────────────────────────────────────────────────┘

            ┌─────────────────────────────────────────────────┐
            │ Pre-processed source document and supporting     │
            │ material made available to the Demand Management  │
            │ Section and translators, monolingual reference    │
            │ corpus expanded                                   │
            └─────────────────────────────────────────────────┘
```

where the 3rd SOP can be described as

```
        ┌─────────────────────────────────────────────┐
        │   Source document pre-processed (automations) │
        └─────────────────────────────────────────────┘
                              │
        ┌─────────────────────────────────────────────┐
        │ 3.1 – Analyse the results of automated        │
        │ pre-processing activities and assign tasks    │
        │ to LTS teams                                  │
        └─────────────────────────────────────────────┘
                              │
        ┌─────────────────────────────────────────────┐
        │ 3.2 – Send source document and results        │
        │ to all LTS teams to be involved               │
        └─────────────────────────────────────────────┘
                              │
        ┌─────────────────────────────────────────────┐
        │   Source document ready to be pre-processed   │
        │        sent to all teams in charge            │
        └─────────────────────────────────────────────┘
```

and the SOPs 4a and 4b as

```
        ┌─────────────────────────────────────────────────────┐
        │  Manual pre-processing request received from the      │
        │                  COORD team                           │
        └─────────────────────────────────────────────────────┘
                │                   │                   │
┌───────────────────────┐ ┌───────────────────┐ ┌───────────────────────┐
│ 4aa – Produce a         │ │ 4ab – Create or   │ │ 4ac – Prepare special │
│ pre-formatted source    │ │ feed translations │ │ format files and      │
│ document which is       │ │ memories          │ │ validate them         │
│ machine readable,       │ │ (CAT team)        │ │ (CAT team)            │
│ editable and which can  │ │                   │ │                       │
│ be used with all tools  │ │                   │ │                       │
│ Made available to       │ │                   │ │                       │
│ translators             │ │                   │ │                       │
│ (COORD, TECH teams)     │ │                   │ │                       │
└───────────────────────┘ └───────────────────┘ └───────────────────────┘
                │                   │                   │
        ┌─────────────────────────────────────────────────────┐
        │  Source document technically pre-processed and        │
        │  results, supporting documents and instructions saved │
        └─────────────────────────────────────────────────────┘


        ┌─────────────────────────────────────────────────────┐
        │  Manual pre-processing request received from the      │
        │                  COORD team                           │
        └─────────────────────────────────────────────────────┘
                │                   │                   │
┌───────────────────────┐ ┌───────────────────┐ ┌───────────────────────┐
│ 4ba – Identify and make │ │ 4bb – Pre-        │ │ 4bc – Do the          │
│ available reference     │ │ translate the     │ │ terminological and    │
│ documents               │ │ source document   │ │ textual mark-up       │
│ (REF team)              │ │ (COORD team)      │ │ (COORD team)          │
└───────────────────────┘ └───────────────────┘ └───────────────────────┘
                │                   │                   │
        ┌─────────────────────────────────────────────────────┐
        │  Source document linguistically pre-processed and     │
        │  results, supporting documents and instructions saved │
        └─────────────────────────────────────────────────────┘
```
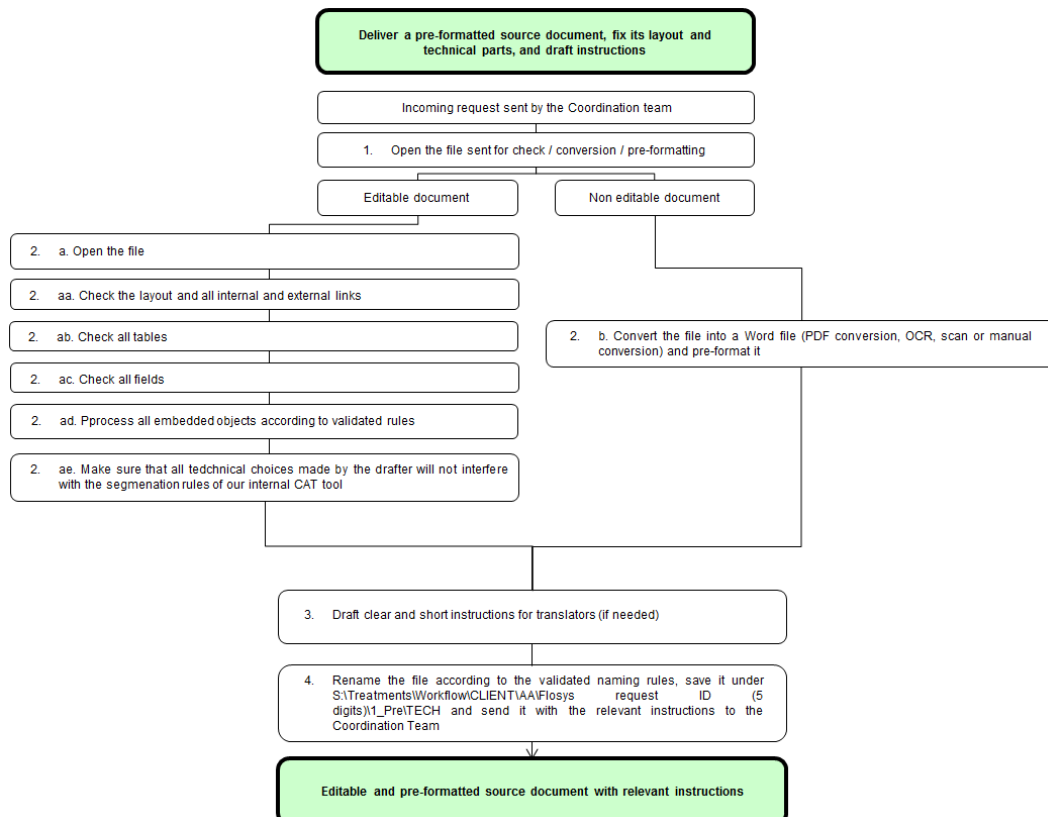
Some 34 detailed work instructions for all possible pre-processing steps designed to cover all identified needs were established:

```
┌─────────────────────────────────────────────┐
│         Pre-translate a source document       │
└─────────────────────────────────────────────┘
┌─────────────────────────────────────────────┐
│ The (pre-formatted, if needed) document is    │
│ available and the Demand Management Section   │
│ needs its pre-translation                     │
└─────────────────────────────────────────────┘
┌─────────────────────────────────────────────┐
│ 1.  Retrieve the outcome of the analyse of    │
│     the source document performed with        │
│     Euramis (vs. CDT and CDT-References)      │
└─────────────────────────────────────────────┘
┌─────────────────────────────────────────────┐
│ Internal translation memories are updated,    │
│ usable and the alignments of the identified   │
│ previous related translations are available   │
└─────────────────────────────────────────────┘
        (Yes)                      (No)
```

Yes branch:

2. a. Pre-translate toe source document into the various target languages listed in the request

2. aa. Clean-up the file and save the RET.doc file as reference material

2. ab. Save all TMX files created with Euramis (CDT)

2. ac. Save th epre-translated document

3. Draft instructions and attach them to the worksheet (Research Assistant Job)

No branch:

2. b. Contact the Reference material Team to get the list of multilingual files available (reference material)e

2. ba. Align available reference material

2. bb. Import alignments in the corresponding translation memories and in Euramis (CDT)

2. bc. Go to 2. a.

```
┌─────────────────────────────────────────────┐
│ Pre-translated source document, TMX files and │
│ instructions made available                    │
└─────────────────────────────────────────────┘
```

and

```
┌─────────────────────────────────────────────┐
│ Deliver a pre-formatted source document, fix  │
│ its layout and technical parts, and draft     │
│ instructions                                   │
└─────────────────────────────────────────────┘
┌─────────────────────────────────────────────┐
│ Incoming request sent by the Coordination team│
└─────────────────────────────────────────────┘
┌─────────────────────────────────────────────┐
│ 1.  Open the file sent for check / conversion │
│     / pre-formatting                          │
└─────────────────────────────────────────────┘
     Editable document    Non editable document
```

Editable document branch:

2. a. Open the file

2. aa. Check the layout and all internal and external links

2. ab. Check all tables

2. ac. Check all fields

2. ad. Pprocess all embedded objects according to validated rules

2. ae. Make sure that all tedchnical choices made by the drafter will not interfere with the segmenation rules of our internal CAT tool

Non editable document branch:

2. b. Convert the file into a Word file (PDF conversion, OCR, scan or manual conversion) and pre-format it

3. Draft clear and short instructions for translators (if needed)

4. Rename the file according to the validated naming rules, save it under S:\Treatments\Workflow\CLIENT\AA\Flosys request ID (5 digits)\1_Pre\TECH and send it with the relevant instructions to the Coordination Team

```
┌─────────────────────────────────────────────┐
│ Editable and pre-formatted source document     │
│ with relevant instructions                     │
└─────────────────────────────────────────────┘
```

Considering that some 60-65% of all processed pages are first handled by freelance translators, a second major challenge emerged: how can all pages processed externally be made workable for in-house linguists and fit for the post-processing (of translations and revisions done in house) stage, i.e. originally for the last quality check before delivery to clients?

Furthermore, building on analyses, remarks and, occasionally, complaints from in-house linguists and clients, another process emerged: the mid-processing procedure (of translations done by freelance translators), i.e. a targeted technical check of files processed by freelance translators before their assignment to in-house linguists.

Consequently, another process had to be defined for the mid-processing procedure, namely a number of activities for checking various elements, particularly in HTML, XML and InDesign files processed externally:
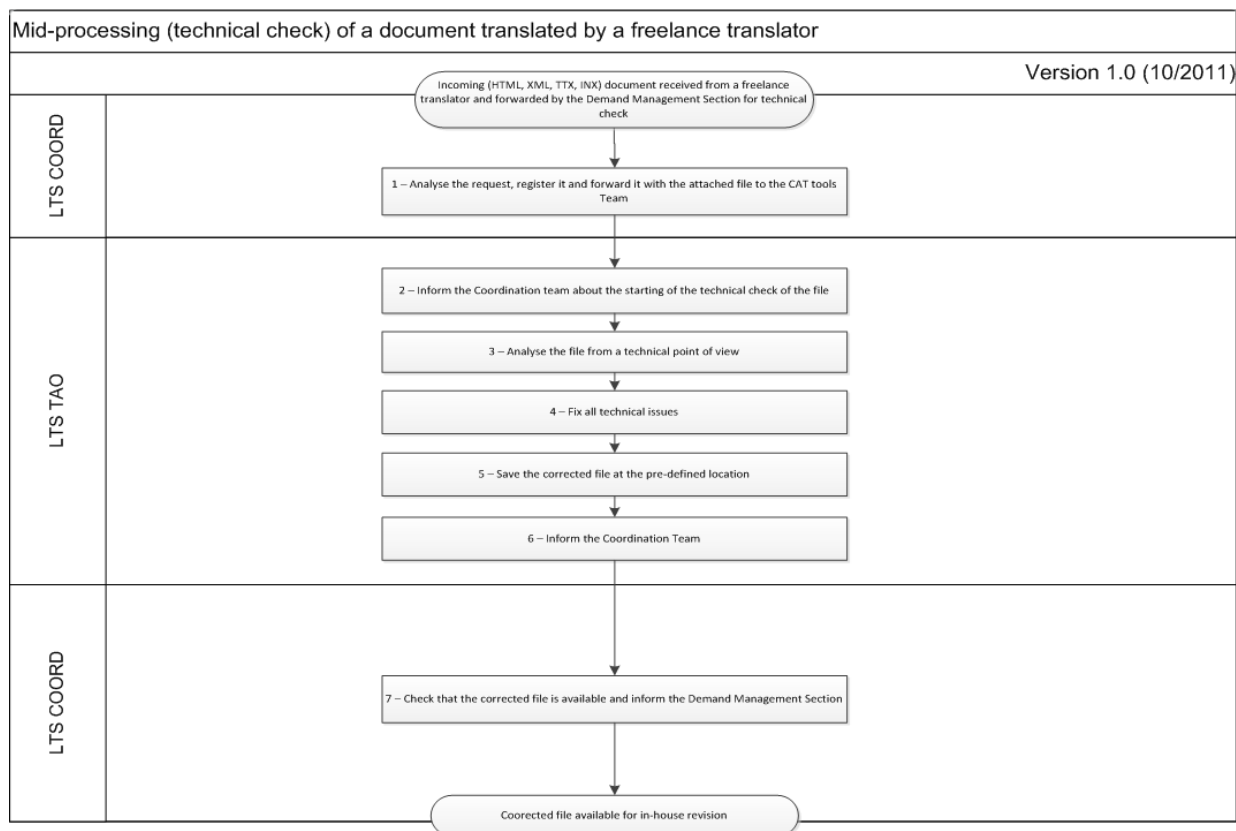
At the same time, some of the work instructions mentioned above where slightly modified to cover the post-formatting and post-alignment activities. LTS then focused on:

- the organisation of the post-processing work (technical check of outgoing files, alignment of documents, terminology extractions);
- the enrichment and creation of resources (documentary fund, translation memories, glossaries) and tools; and
- the introduction of a new technical step concerning translations done by freelancers that was necessary before the documents were assigned to in-house linguists in order to avoid recurring technical issues identified by clients or the Technical team.

Basic management tools such as workflows, processes, indicators and workload and productivity measurement were defined and implemented, and are updated whenever necessary. Thanks to a diversification of approaches, a melting pot of ideas and a growing conception and identification of tools, as well as total openness to in-house colleagues and the use of various reporting channels, i.e. building on open-door and brainstorming sessions with various in-house and external stakeholders and on their considerable feedback, all these processes are kept alive and evolve according to needs identified and to any emerging weaknesses relating to new expectations or to technological developments. Furthermore, LTS is dedicated to improving its

results and clearly demonstrates its strong will to remain flexible for addressing new needs and facing all challenges.

A further building block of the 'revolution' launched has been the focus placed on skills since the end of 2010. Building on internal and external training sessions and new insight for all team members in all activities covered by all teams, multitasking and multi-skilling slowly became a reality, allowing members of all teams to better understand the work performed by others and to perform some basic activities in other fields. This trend is to be linked with some further positive side-effects:

- organisation of in-house training sessions for LTS staff members;
- organisation of in-house training sessions for linguists;
- organisation of training sessions for authors, secretaries and all staff members interested;
- improved skills for all LTS staff members providing a helpdesk service (on request) to linguists (for CAT, formatting, reference material and terminology issues);
- greater collaboration with other departments and sections (Translation and IT Departments, External Relations and Communication Section and Human Resources Section);
- hosting of trainees (in the terminology and CAT tools fields), bringing knowledge and new ideas to be transferred into in-house IT developments, new tools and new resources;
- creation of tools for in-house use, including for linguists;
- improved networking with external peers and partners.

Symptomatically, while mid-processing remains an activity without any immediate visibility, pre- and post-processing activities are now requested by some clients as services as such, namely:

- conversion of non-editable documents (e.g. in PDF format) into Word files, while applying predefined formatting rules;
- alignment of documents and TMX file generation with insertion of predefined attributes;
- reformatting and layout of files according to specific rules.

While all questions listed above were answered with the definition of three main processes and their regular review, another came to light:

> What further measures can be taken 'behind the scenes' to make linguists' jobs easier?

Focusing on linguists' needs in terms of tools and resources, a further process has been considered, and is currently under definition, to formalise ways identified for enriching TMX corpora and all available document and terminology resources.

For example, we have already defined a related SOP, focused on the building of new multilingual terminology resources:
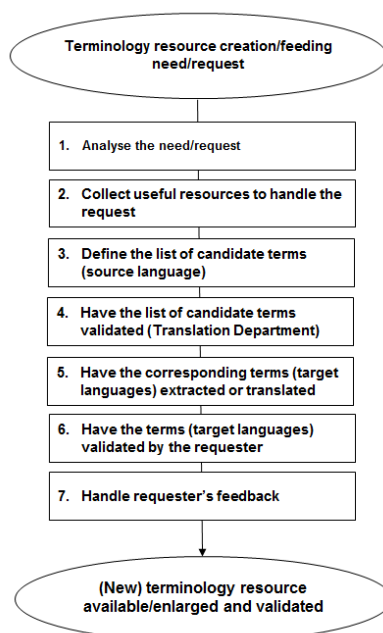
Looking back over the past three years:

- The lines between specialised teams are becoming increasingly hazy, thus opening and broadening possibilities and approaches, as skills and know-how are becoming increasingly mixed;
- LTS has been built on a solid basis of knowledge and skills, in which most staff are multitasked and multiskilled and mindful of technological changes;
- LTS took on board a systematic pre-processing of all incoming requests and mid-processing activities for files in specific defined formats, and enlarged post-processing activities to strengthen and enrich resources made (directly or indirectly) available to in-house and freelance linguists;
- Some 16 staff members now provide language and technical support to hundreds of freelance linguists;
- Those 16 staff members provide the same support to all in-house linguists, despite the increased number of official EU languages;
- LTS now handles freelance requests, making it a direct service provider for the Centre's freelance partners.
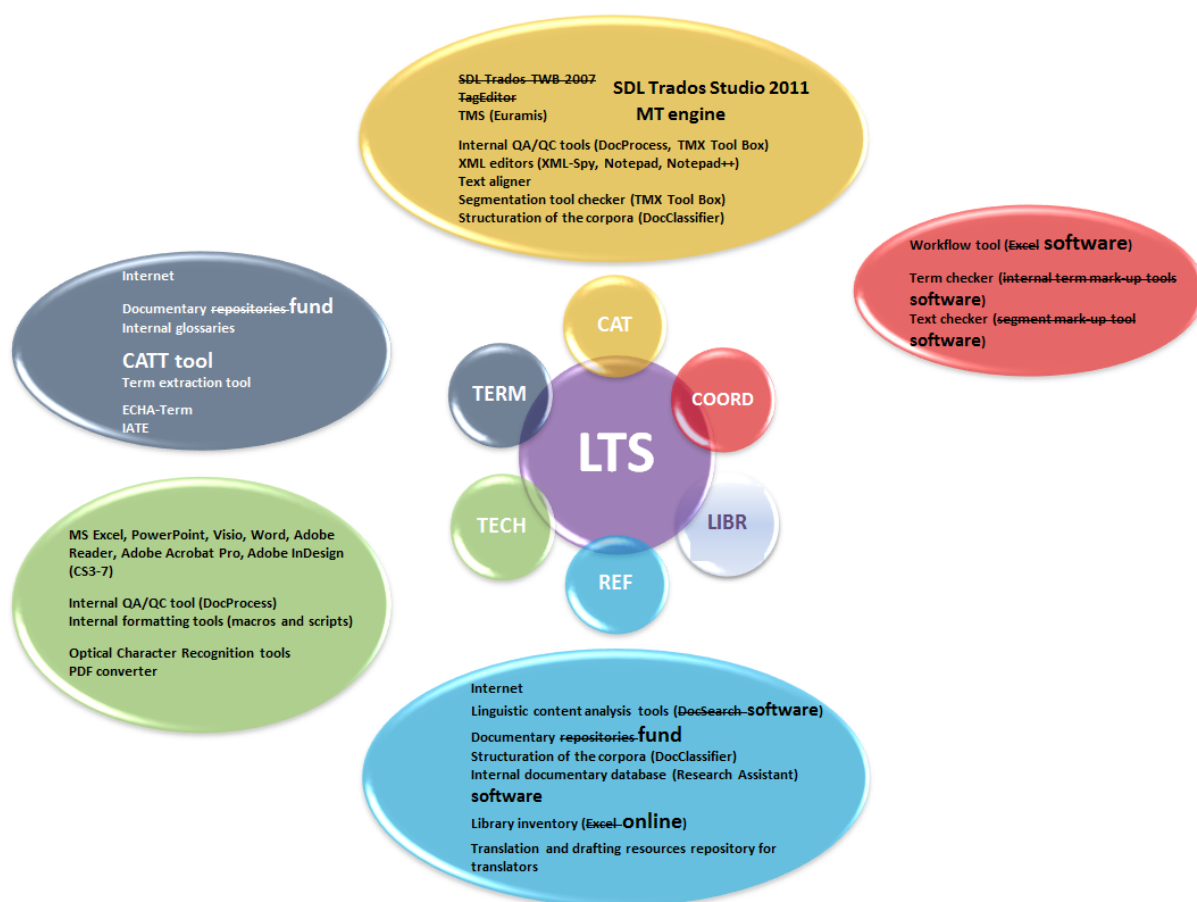
In terms of (pre-, mid- and post-processing) results, we note the following:

	2012*	01-10/2013**
Available staff (FTE) for document (pre-, mid-, post-) processing activities	10.8	9.7
Available staff for document (pre-, mid-, post-) processing activities (%)	78.4%	74.6%
Pre-processing		
Number of incoming source documents		
Number of incoming pages	89 128	109 114
Variation vs previous year (same month)		
% of previous year (cumulated)	+110.8%	+22.4%
Number of incoming Word pages	77 558	96 373
Variation vs previous year (same month)		
% of previous year (cumulated)	+3.9%	+24.3%
Number of incoming non-Word pages	11 570	12 741
Variation vs previous year (same month)		
% of previous year (cumulated)	+81%	+10.1%
% of pre-processed incoming documents	100.0%	100.0%
Post-processing		
Number of outgoing pages	329 132	260 931
Variation vs previous year (same month)		
% of previous year (cumulated)	+152.2%	-20.7%
Number of outgoing Word pages	289 400	227 334
Variation vs previous year (same month)		
% of previous year (cumulated)	+141.1%	-21.4%
Number of outgoing non-Word pages	39 732	33 597
Variation vs previous year (same month)		
% of previous year (cumulated)	+356.9%	-15.4%
Maintenance		
Number of pages aligned	157 007	99 593
Quality		
% of TMs with technical incidents	0.38%	0.08%
% of justified negative feedback from translators	0.20%	0.13%
% of justified negative feedback from translators (LING)	0.02%	0.02%
% of justified negative feedback from translators (TECH)	0.24%	0.12%
% of justified negative CSFs	0.07%	0.01%
	* 12 months	** 10 months

What next? The original five teams will merge to some extent and a new common approach should lead to the further, precise structuring of the vast corpora of documents we processed for our clients over the past 15 years. But what are the (new) challenges to be faced? The implementation of new (CAT, CATT and corpus management) tools will further contribute to the strengthening of synergies between teams and will pave the way for the introduction and integration of machine translation into the translation workflow.

As a result, we will have a switch in the tools used for all processing activities:

SDL Trados TWB 2007
TagEditor
TMS (Euramis)

**SDL Trados Studio 2011
MT engine**

Internal QA/QC tools (DocProcess, TMX Tool Box)
XML editors (XML-Spy, Notepad, Notepad++)
Text aligner
Segmentation tool checker (TMX Tool Box)
Structuration of the corpora (DocClassifier)

Workflow tool (Excel **software**)

Term checker (internal term mark-up tools **software**)
Text checker (segment mark-up tool **software**)

Internet
Documentary repositories **fund**
Internal glossaries
CATT tool
Term extraction tool
ECHA-Term
IATE

MS Excel, PowerPoint, Visio, Word, Adobe Reader, Adobe Acrobat Pro, Adobe InDesign (CS3-7)
Internal QA/QC tool (DocProcess)
Internal formatting tools (macros and scripts)
Optical Character Recognition tools
PDF converter

CAT
TERM
COORD
LTS
TECH
LIBR
REF

Internet
Linguistic content analysis tools (DocSearch **software**)
Documentary repositories **fund**
Structuration of the corpora (DocClassifier)
Internal documentary database (Research Assistant) **software**
Library inventory (Excel **online**)
Translation and drafting resources repository for translators

Therefore, we are reorganising several activities and prepare the implementation of all new tools, which means e.g.:

- drafting training material in view of the implementation of the new tools,
- training staff members,
- redefining linguists' needs in view of the new tools' functionalities,
- rethinking our work organisation,
- redocumenting LTS' processes, SOPs and WINs,
- further structuring our corpora (definition of new metadata/attributes, identification of keywords, use of an internal tool (DocClassifier),
- further integrating clients' new needs in the pre-, mid- and post-processing stages (e.g. expanded localisation work and the related Website content management in some cases),
- exploring possibilities in new fields (subtitling, machine translation, e.g.).

Finally yet importantly, as no formatting rules have been defined and agreed, clients may send documents in any format meeting their needs, which means that LTS has to deal with – and will have to continue dealing with in the future – all kinds of formats used by clients.

In a nutshell: analyse and structure your working methods and keep them alive, build bridges, focus on novelty, promote communication and knowledge exchange to improve and expand background activities on a continuous basis in order to speed up the translation process and make clients' satisfaction central to your list of values!

---

[i] In this document, 'linguists' means 'in-house translators, revisers, editors and freelance translators and editors'.