

So you have term base functionality. Now what?

Best practices in managing
terminology and term bases in the
MemoQ translation environment

© 2009 Douglas McCarthy



Extending the concept of terminological consistency/coherence

using the concept of lexical cohesion

Example: the MemoQ documentation



An excerpt from the MemoQ documentation

Overview	A term base is a database containing pairs of source expressions and their translations. Term bases are used to maintain translation consistency and simplify the translation of demanding projects with specific terminology.
Types Of Term Bases	Depending on their physical location, MemoQ recognizes three types of term bases: <ul style="list-style-type: none">• <i>Local term base</i> is physically present on a user's hard drive.• <i>Remote term base</i> is located in a remote computer and can be accessed through the Internet.• <i>Synchronized remote term base</i> is a hybrid of the above two. The primary copy is located in a remote computer and can be accessed through the Internet, however, there is a copy stored on your computer, and you can also work with the term base when you are not connected to the Internet for any reason. When you connect to the Internet again, the remote and the local term bases are synchronized, i.e. the remote copy is updated with your new terms added offline, and the local copy is updated with the changes in the contents of the remote term base.



Lexical cohesion

Collocations

term base	primary copy
local term base	remote copy
remote term base	local copy
synchronized base term	update ... copy
primary term base	primary copy
moderated term base	suggest new term
unmoderated term base	add new term
	Register remote command
	Synchronize offline command
	suggest new term

+ database = a lexical set.



A 2nd excerpt from the MemoQ documentation

You can also edit remote term bases if you have sufficient privileges. However, one term base may only be edited by one person at a time. Because of heavy network traffic, editing remote term bases may be slow, especially when the server has a slow outbound connection



Let's change the text as follows:

If the user has read and write privileges, remote term bases may also be edited. However, only one term base may be edited by a single user at any given time.

Networks may experience latency due to heavy traffic. This can increase editing time significantly, particularly when it is the download connection with the server that is characterized by latency.



What's the difference?

Register and genre are different: the second text has the same content but uses terms appropriate for describing the technology to engineers rather than end-users. At the level of lexis, note: *latency, read and write privileges*



Two translation examples where the translator didn't take this type of feature into account



Poor copy

In reality, Paris was just a start for you, and when your father passed you the reigns in 1988, the company grew unremittingly: now, in over 40 countries, more than 3,500 people work for the company and you develop products in several fields, of which my hormonal connection and my independence allow me to stress their very great quality.



Source text

Au fond, Paris n'est qu'un début pour vous et alors que votre père vous passe la main en 1988, la société ne cesse de grandir : maintenant dans plus de 40 pays, plus de 3500 personnes y travaillent et vous développez des produits dans plusieurs domaines dont la connotation hormonale et mon indépendance me permettent de souligner la très grande qualité.



Corrected copy

In reality, Paris was just a start for you, and when your father passed you the reins in 1988, the company grew unremittingly: it is now present in over 40 countries; more than 3,500 people work for it, and you develop products in several fields. On the basis of the hormonal aspect of one of these fields, and my own independence in this area, I am well placed to emphasise their very high quality.



Poor copy

22 LSB/USB Button

- To select the lower or upper side band on **SSB** (Single Side Band) reception.

Stepping button, steps forward in the order:

With reception frequency < 10 MHz = LSB \rightarrow USB \rightarrow Normal \rightarrow LSB, etc.

With reception frequency ≥ 10 MHz = USB \rightarrow LSB \rightarrow Normal \rightarrow USB, etc.



Source text

22 LSB/USB-Taste

Einstellen des unteren bzw. Oberen Seitenbandes bei
SSB (Single-Side-Band)-Empfang.

Fortschalt-Taste, schaltet weiter in der Reihenfolge:

Bei Empfangsfrequenz < 10 MHz = LSB \rightarrow USB \rightarrow Normal
 \rightarrow LSB usw.

Bei Empfangsfrequenz ≥ 10 MHz = USB \rightarrow LSB \rightarrow Normal
 \rightarrow USB usw.



Corrected copy

22 LSB/USB Button

- Selecting the lower or upper side band on **SSB** (Single-sideband) reception.

Every time the step switch button is pressed, the selection changes in the following order:

With receiving frequency < 10 MHz = LSB \rightarrow USB \rightarrow Normal \rightarrow LSB, etc.

With receiving frequency ≥ 10 MHz = USB \rightarrow LSB \rightarrow Normal \rightarrow USB, etc.



Other lexical features that can be dealt with in a term base

Multiple mappings of a single word in a single text due to context, stylistic variation, error:

- Treasury shares/Own shares
- chairman of committee, president of committee



Terminological

coherence/terminological consistency

Concept needs to be extended because it doesn't take account of:

- Multiple mappings of a single word in a single text on the basis of context
- Role of genre, register and mode
- No emphasis on information such as grammar



- Changes in client language policy over time
 - Choice of terms can change from one version to another (e.g. in source language in v.1 and in English in v.2: *Écart d'acquisition* -> *goodwill*)
- The time factor in the framework of the project
 - Term translations may change status from *proposed* to *confirmed* to *validated*



Practically speaking, we can already implement these ideas using MemoQ

1. Converting client glossaries into term bases
2. Enriching our termbases by taking advantage of some of the fields available
3. Communicating with our client by exporting the termbase as a dictionary for client validation



Legacy glossaries may need to be taken out of a document with broader scope

The screenshot shows the OpenOffice.org Writer interface. The main document window displays a text document with the following content:

It is essential, to make everybody's task easier that you stick to this terminology.
If, however you do have some questions/concerns with regard to any of the following words please let us know **IN ADVANCE** in order for us to set a common guideline for all translators.

Also, we insist that you send us any new terminology questions **BEFORE** giving in your translations.

PLEASE TAKE THE TIME TO READ THESE FEW GUIDELINES BEFORE YOU START AND NOT ONLY IF YOU'RE STUCK!
We are fully aware that you are all extremely efficient as far as terminology is concerned we would just like to make sure everyone use the same terms.

Abbreviations/acronyms

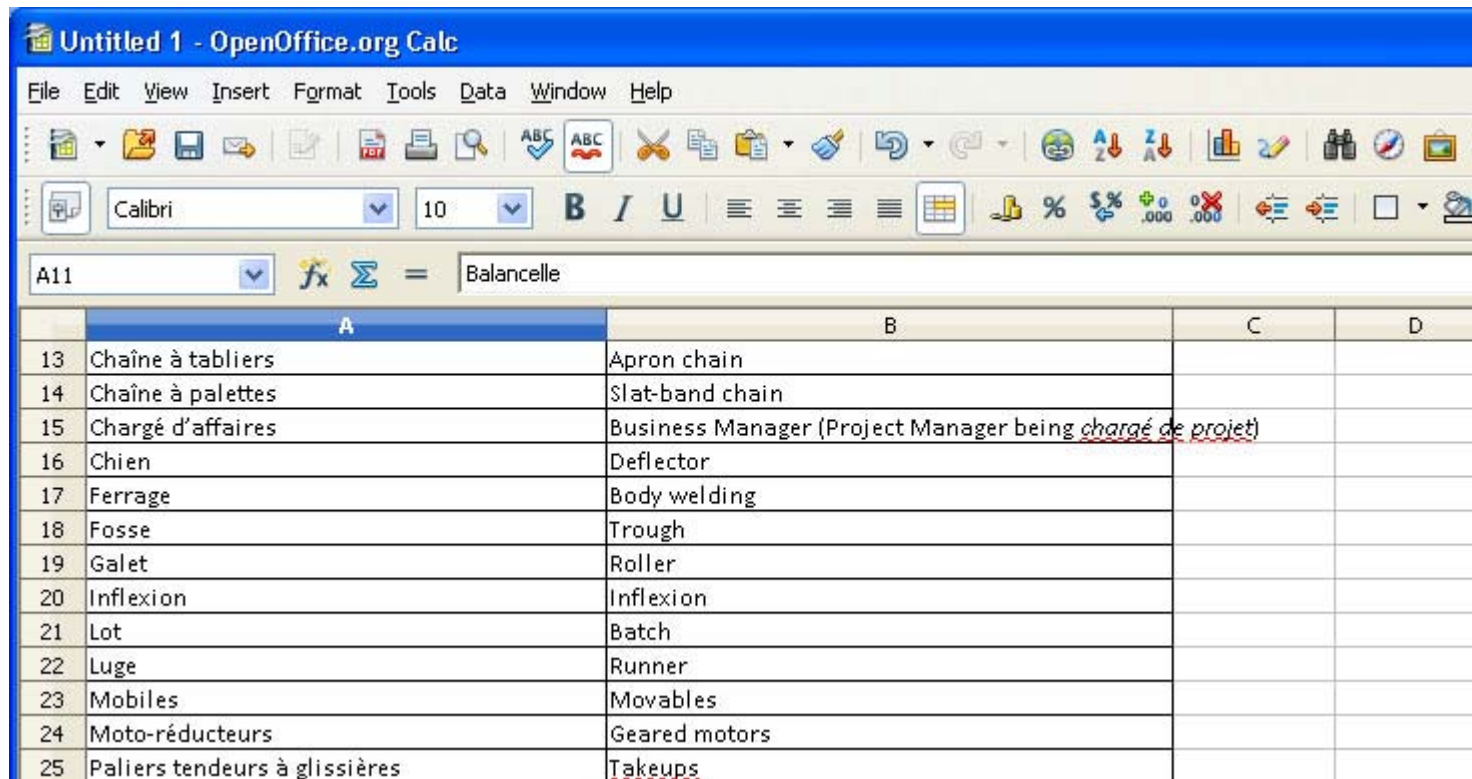
Source language	Target language
AMD EC - Analyse des Modes de Défaillance, de leurs Effets et de leur Criticité	FMECA - Failure Mode, Effects, and Criticality Analysis
API	API
AU et AF	ES and PS
DA - Direction Achats	PD - Purchasing Department
GEST NOM	GEST NOM
HEB	This refers to a type of beam in the shape of a "H" so

On the right side of the interface, a list of styles is visible, including: Addressee, Caption, Drawing, Endnote, Footer, Footer left, Footer right, Footnote, Frame contents, Header, Header left, Header right, Illustration, Sender, Table, Table Contents, Table Heading, and Text.

g up dial-up connection when all ...
mnection.com/titles/media16.htm - 95k - [Cached](#) - [Similar pages](#) -



Paste the relevant table into a blank spreadsheet



The screenshot shows a spreadsheet window titled "Untitled 1 - OpenOffice.org Calc". The spreadsheet contains a table with 13 rows of data. The first column (A) contains French terms, and the second column (B) contains their English equivalents. The third and fourth columns (C and D) are empty.

	A	B	C	D
13	Chaîne à tabliers	Apron chain		
14	Chaîne à palettes	Slat-band chain		
15	Chargé d'affaires	Business Manager (Project Manager being <i>chargé de projet</i>)		
16	Chien	Deflector		
17	Ferrage	Body welding		
18	Fosse	Trough		
19	Galet	Roller		
20	Inflexion	Inflexion		
21	Lot	Batch		
22	Luge	Runner		
23	Mobiles	Movables		
24	Moto-réducteurs	Geared motors		
25	Paliers tendeurs à glissières	Takeups		

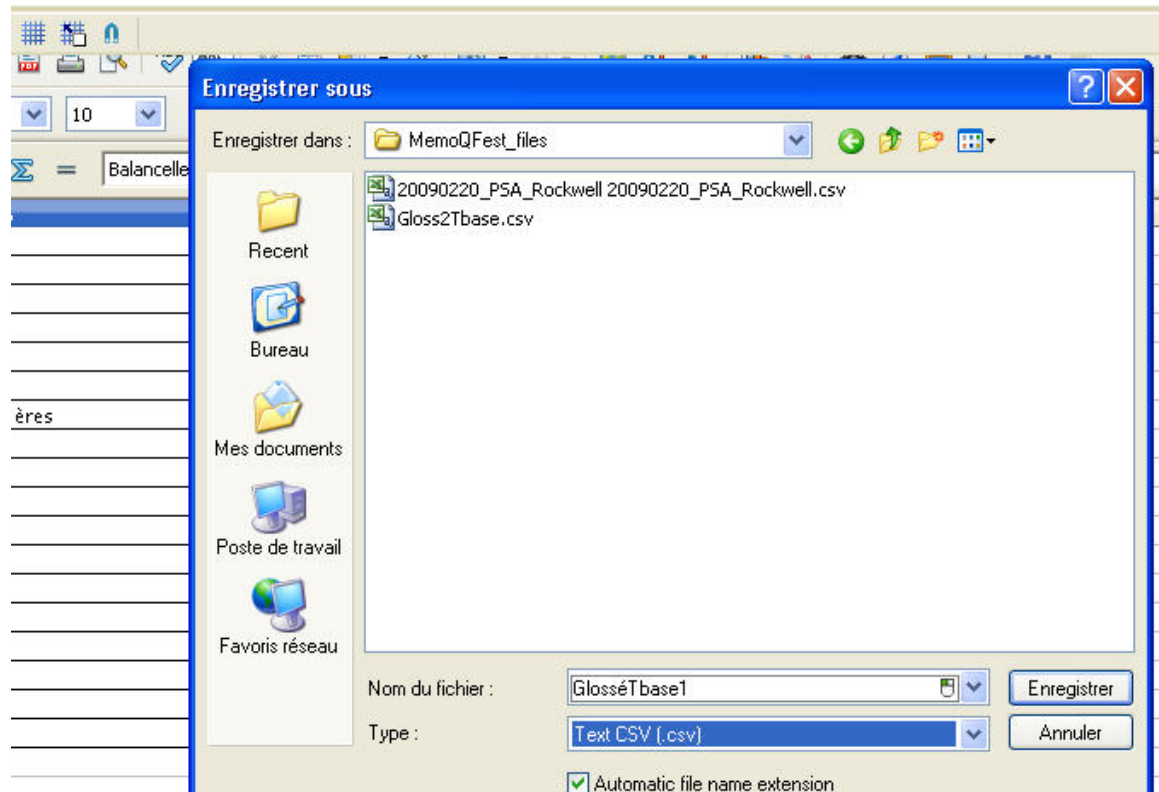


Take out cells striding over 2 columns
 (General Terminology below)

4	DA – Direction Achats	PD – Purchasing Department	
5	GESTNOM	GESTNOM	
6	HEB	This refers to a type of beam in the shape of a “H” so “H-Beam” or H-shape	
7	MEF	Production Run	
8	PEE – Procédures d’Exécution d’Essais	TEP – Test Execution Procedures	
9	TC – Temps de Cycle	CT – Cycle Time	
10	TOR – Tout Ou Rien	AON - All Or Nothing	
11	General Terminology		
12	Balancelle	Swing tray	
13	Cahier des charges (CDC)	Specifications (specs)	
14	Chaîne à tabliers	Apron chain	
15	Chaîne à palettes	Slat-band chain	
16	Chargé d’affaires	Business Manager (Project Manager being <i>chargé de projet</i>)	
17	Chien	Reflector	



Save the spreadsheet as a CSV (comma separated value) file



And import the CSV file into the MemoQ term base

The screenshot shows the 'Term bases' dialog box in MemoQ. On the left is a sidebar with icons for Translations, Alignments, Translation memories, Term bases (highlighted), and Settings. The main area is divided into two tables: 'Term bases in project' and 'Available term bases (filtered)'. Both tables have columns for Name, Location, and Path/URL. In the 'Term bases in project' table, 'demo' is selected. In the 'Available term bases (filtered)' table, 'demo' is highlighted. Below the tables are several action buttons: 'Set as primary', '<< Add selected', '>> Remove selected', 'Create/add new local', 'Register remote', 'Register local', 'Export', 'Properties', 'Delete', 'Edit', 'Import from CSV/TMX', and 'Repair'. A mouse cursor is pointing at the 'Import from CSV/TMX' button.

Name	Location	Path/URL
20090501_Drucker1	Local	C:\Documents and ...
demo	Local	C:\Documents and ...

Name	Location	Path/URL
20090218_D\TheClient ...	Local	C:\Documents and ...
20090219_Renault 20090...	Local	C:\Documents and ...
20090224_Tagline_GE 20...	Local	C:\Documents and ...
20090501_Drucker	Local	C:\Documents and ...
20090501_Drucker1	Local	C:\Documents and ...
demo	Local	C:\Documents and ...



- Enriching term bases:
 - By adding images to fix the difference between terms in the translator's mind

ID	Source	Target	Author	Date
10	Balanceuse	swing tray	dmy	20/04/2009
11	Cahier des charges (CDC)	Specifications (specs)	dmy	20/04/2009
12	Chaîne à tabliers	Apron chain	dmy	20/04/2009
13	Chaîne à palettes	Slat-band chain	dmy	20/04/2009
14	Chargé d'affaires	Business Manager (Project Manager being chargé de projet)	dmy	20/04/2009
15	Chien	Deflector	dmy	20/04/2009
16	Ferrage	Body welding	dmy	20/04/2009
17	Fosse	Trough	dmy	20/04/2009
18	Galot	Roller	dmy	20/04/2009

The screenshot shows the memoQ interface for editing a term. On the left, the 'French (France)' column contains 'Chaîne à tabliers' and the 'English (United Kingdom)' column contains 'Apron chain'. Below the text fields are tabs for 'Matching', 'Usage', 'Grammar', and 'Definition'. The 'Matching' tab is active, showing 'Prefix matching' set to '50% minimum' and 'Case sensitivity' set to 'Permissive'. On the right, there is an 'Image' tab with buttons for 'Import new image', 'Delete image', and 'Show full-size image'. A checkbox labeled 'Show images' is checked. A large image of a blue industrial apron chain is displayed in a window on the right.



– And to differentiate similar terms

	Apron chain	dmy	05/05/20
	Slat-band chain	dmy	20/04/20
	Business Manager (Project Manager being chargé de projet)	dmy	20/04/20
	Deflector	dmy	20/04/20
	Body welding	dmy	20/04/20
	Trough	dmy	20/04/20
	Roller	dmy	20/04/20

The screenshot shows the memoQ software interface. The language is set to English (United Kingdom). The term 'Slat-band chain' is selected in the list. The 'Image' tab is active, showing a small image of a slat-band chain. The interface includes buttons for 'Import new image', 'Delete image', and 'Show full-size image'. There is also a checkbox for 'Show images' which is checked. The 'Matching' tab is also visible, showing 'Prefix matching' set to '50% minimum' and 'Case sensitivity' set to 'Permissive'.



- Use word forms that are actually found in the text, not canonical forms (i.e. infinitives and singular, nominative)



Using the term base to communicate with the client

- From a CSV export to a spreadsheet (Excel or OpenOffice.org, for example)
- From a MultiTerm XML export to an RTF dictionary (via MultiTerm for the time being*)

*2 wish list items here:

1. An XSLT template to convert the XML file to an RTF dictionary format directly without passing through MultiTerm
2. The addition of a status field in MemoQ term bases to map proposed statuses directly into the corresponding field in MultiTerm format



Export to MultiTerm ... in brief

Term bases

Translations

Alignments

Translation memories

Term bases

Settings

Term bases in project

Name	Location	Path/URL
✓ 20090501_Drucker1	Local	C:\Documents and ...
✓ demo	Local	C:\Documents and ...

Available

Name	Location
✓ 20090218_ZX_ThinClient ...	Local
✓ 20090219_Renault 20090...	Local
✓ 20090224_Tagline_GE 20...	Local
✓ 20090501_Drucker	Local
✓ 20090501_Drucker1	Local
✓ demo	Local

• Set as primary

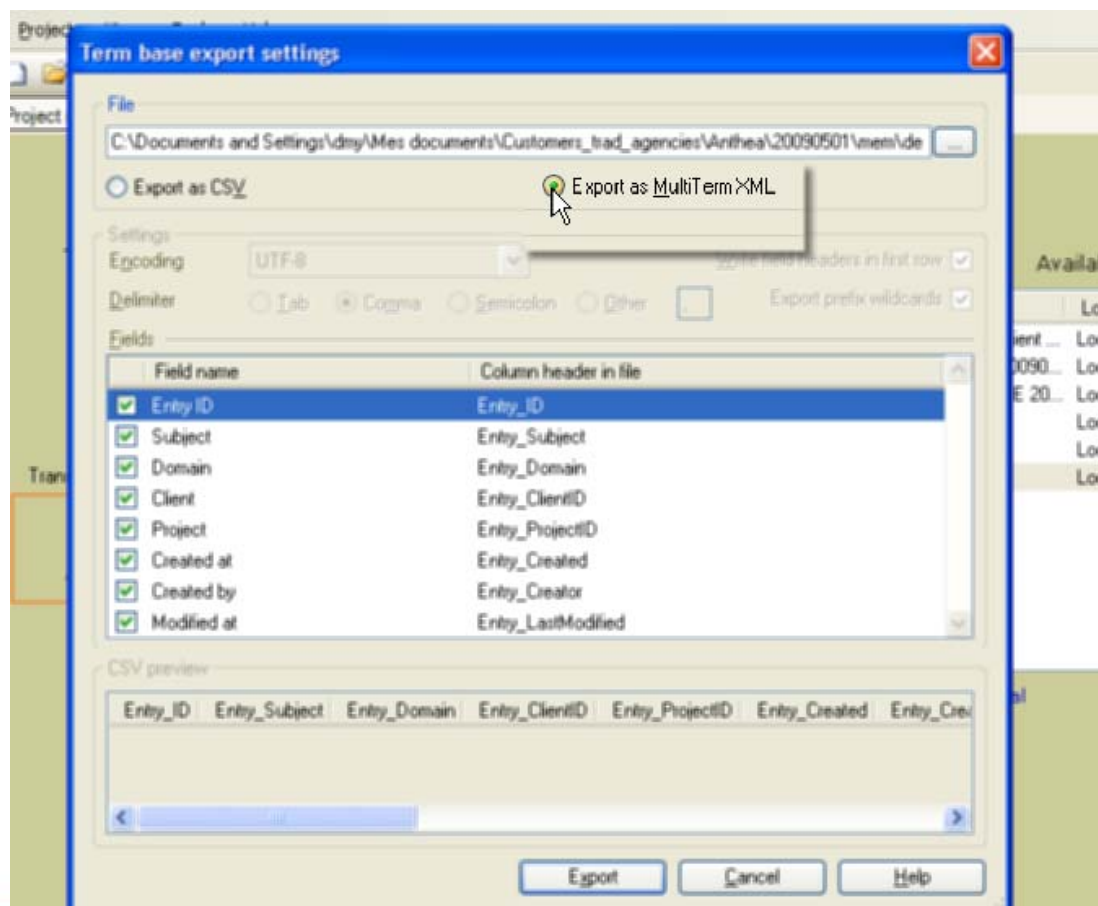
<< Add selected •

>> Remove selected •

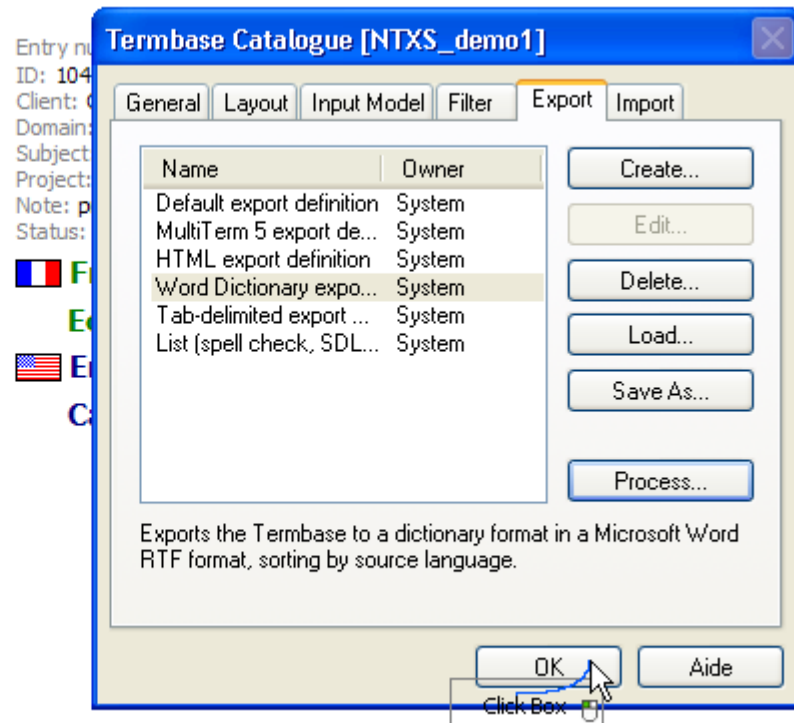
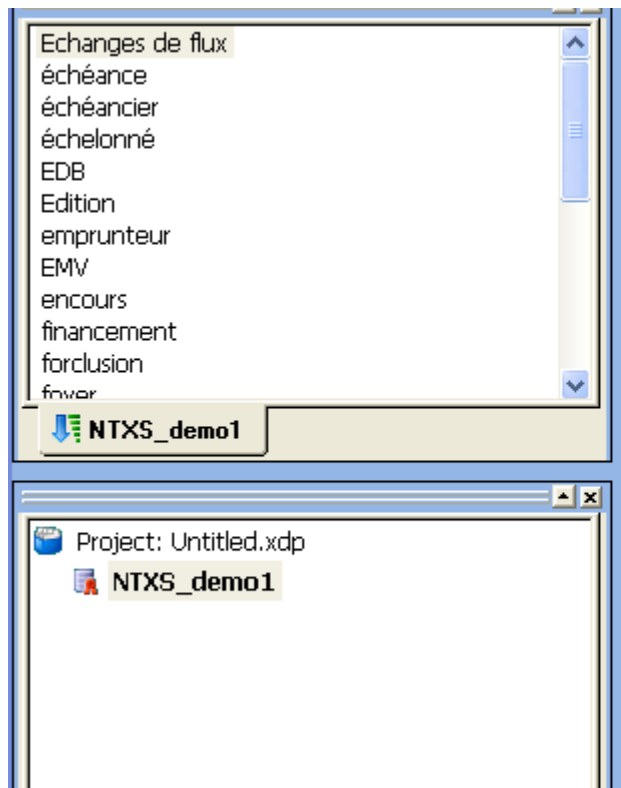
- Create/add new local
- Register remote
- Register local
- Export
- Properties



Choosing between CSV and MultiTerm XML



Export from MultiTerm to Word dictionary



The dictionary output

English (United States) Disbursement

DMA ID 27 Domain Banking Subject Insurance
English (United States) **DMA**

DOM ID 28 Domain Banking Subject Insurance
English (United States) **DOM**

DTS ID 29 Domain Banking Subject Insurance
English (United States) **DTS**

E

Echanges de flux Subject Insurance Note
proposed Status proposed ID 104 Client G...
Domain BANKING Project 20081 208
English (United States) **Cash flow swaps**

é

échéance ID 30 Domain Banking Subject
Insurance
English (United States) **Instalment**

English (United States) Disbursement

forclusion ID 38 Domain Banking Subject
Insurance

foyer ID 39 Domain Banking Subject Insurance

FTXTEL ID 40 Domain Banking Subject
Insurance
English (United States) **FTXTEL**

G

GBP ID 41 Domain Banking Subject Insurance
English (United States) **GBP**

GC ID 42 Domain Banking Subject Insurance
English (United States) **GC**

GCE ID 43 Domain Banking Subject Insurance
English (United States) **GCE**

GRC ID 44 Domain Banking Subject Insurance



Best Practices



1. Import legacy glossaries into the translation environment as term bases.
 - a. Select tabular data only
 - b. Check character set encoding & separators
 - c. Import at project creation time
 - d. Use the term base editor to adapt
 - e. Aim for 1 to 1 matching



2. Enrich imported legacy data

- a. Do a preliminary search of source document
- b. Identify key terms and features not in legacy material
- c. Extend the term base with proposed new term translations
- d. Use wild cards if appropriate
- e. Use actual word forms found in text



3. If you're managing a team project on a MemoQ Server ...
 - a) Exploit permissions
 - b) Encourage use of the term base by all team members
 - c) Extend the *terminologist's* role



4. Make proposals for extensions and changes on the basis of:
 - a) Lexical collocations and sets in the document
 - b) Lexical collocations and sets in comparable documents from previous projects or found in the target language on the Internet
 - c) Clearest lexical features that mark register, genre and mode (found in your research)



5. Validate proposals with the client
 - a. Exporting data to CSV or Multiterm
 - b. If you don't have Multiterm, use this data to create a spreadsheet dictionary
 - c. If you have Multiterm, use the RTF dictionary template
 - d. In both cases, specify new term status clearly



6. Exploit QA feature

- a) In individual translator proofreading
- b) In the actual proofreader phase
- c) As the basis of a report to the client on any inconsistent terminology in the document



Conclusion

- Terminological coherence/consistency needs to be extended by lexical cohesion
- Best practices must be introduced, adapted and **reviewed**
- Translation will no longer be the black box in publishing corporate documents

Prepare for the future now!

