

## DEALING WITH TAGS

Tuomas Kostiainen (tuomas@jps.net)

[www.finntranslations.com](http://www.finntranslations.com)

---

**Abstract:** The move from Word as the translation environment to CAT tool-specific environments, such as Trados Studio and memoQ, has created a new problem for many translators: tags. What are they, where do they come from and do we really need them? This presentation answers those questions and shows how to deal with tags during translation and how to avoid them in the first place by using proper settings or external tag-cleaning tools, such as CodeZapper. The CAT tools used during the presentation are Trados Studio, memoQ and Wordfast but the information is also useful for users of other tools.

### 1. What are tags?

Tags contain information about document formatting (such as bold, italics) and structure (such as paragraph styles, tables) and also function as placeholders (for graphics, bookmarks, etc.).

- Inline tags vs. structure tags
- Paired tags (tag pairs) vs. stand-alone tags (placeholder tags)
- Opening tags vs. closing tags
- Ghost tags
- Translatable tags vs. non-translatable tags

### 2. Where do they come from?

Tags come from the source document:

- Formatting, bookmarks, footnotes, placeholders for graphics, etc.

### 3. What's the problem?

Tags can be problematic during translation for various reasons:

- Visual effect
- Translation gets slower and more complicated
- Easier to miss "necessary" tags
- Unnecessary QA warnings
- Effect on match values and TM quality
- Problems when saving target files

### 4. Solutions

- Minimizing the number of tags in source files – prep source files
- Understanding tags – learn to know tags and understand their functions
- Knowing how to use the tag-related settings/features in your CAT tool
- Using tags selectively in target files – not all tags need to be in the target file

#### 4.1 Preparing tag-rich Word files for translation

- Manually in Word
  - Complete cleanup: Recycle the file via a text editor
  - Clearing all text level formatting: Ctrl+Space bar
  - Resetting character spacing, kerning, font, etc.: Ctrl+A, Ctrl+D
  - Deleting bookmarks: Insert > Bookmark > Delete
  - Deleting all hyperlinks: Ctrl+A, Ctrl+Shift+F9
- Clean the file (DOCX, DOC or RTF) in Word using [CodeZapper](#)
  - Use functions CZL, PDFFix and PDFTidy to clean converted PDF files
- Clean the file (DOCX, DOC or RTF) in Word using [TransTools](#) Document Cleaner
  - **Reformat** command performs the following (optional) operations to minimize tags and make the document more user-friendly (see Figure 1):
    - removes text and paragraph shading
    - removes text highlighting
    - resets uneven character spacing
    - removes manual hyphenation
    - fixes formatting problems
    - turns “black” font color into “automatic” color
  - **Bookmark Cleanup** removes specific types of bookmarks, such as bookmarks which are not referenced from fields or hyperlinks, or “table of contents” bookmarks, etc. (see Figure 2)

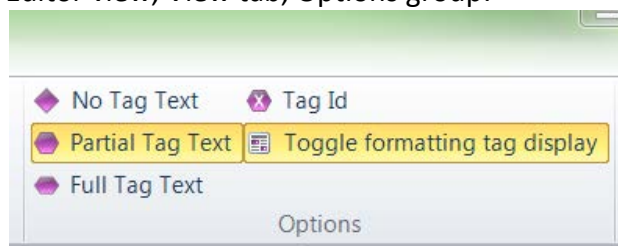
#### 4.2 Using the correct conversion/import settings in your CAT tool

- Trados Studio
  - DOCX: File Types > Microsoft Word 2007-2013 > Common > “[Skip advanced font formatting](#)” (see Figure 3)
- MemoQ
  - DOCX: [Document import settings](#) > Fidelity > “Ignore minor formatting changes for fewer tags” (see Figure 4)
- WordFast
  - No related settings

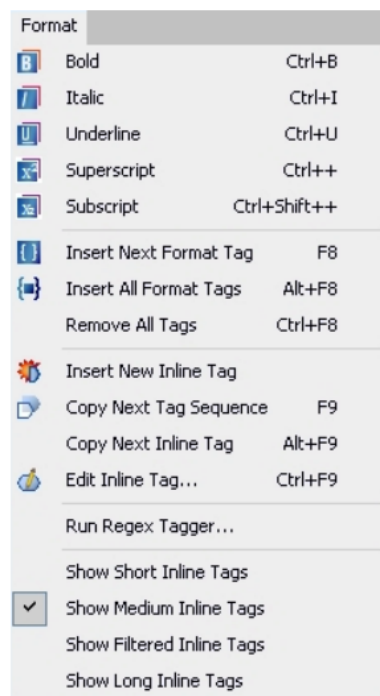
#### 4.3 Selecting the most practical tag display settings in your CAT tool

- Trados Studio
  - For general information about working with tags in Trados Studio, see this [blog article](#) and the [Studio Help](#)
  - No/Partial/Full Tag Text or Tag Id
  - Toggle formatting tag display (on/off)
  - Show all tags but do not show formatting (Options > Editor > Formatting display style)
  - You can also view external tags (Display filter: All content)

- Editor view, View tab, Options group:



- memoQ
  - Tags are not used for basic formatting (bold, italics, underline)
  - [Format](#) menu: Show Short/Medium/Filtered/Long Inline Tags



- WordFast
  - Tag format: {ut1}
  - No tag display settings

#### 4.4 Inserting tags into the target text

- Trados Studio
  - Several methods: Insert individual tags or tag pairs (formatting) from source [with Ctrl+click](#), insert individual tags from the [QuickPlace list](#) [Ctrl+,] or from the Insert Tag dialog box, Ctrl+Alt+right/left arrow, type <, copy & paste, Copy Source to Target [Ctrl+Ins]
- memoQ
  - Tags in DOC (XLS, PPT) files are “[uninterpreted formatting tags](#)”. They cannot be omitted in the target and their order cannot be changed. For DOCX (XLSX, PPTX)

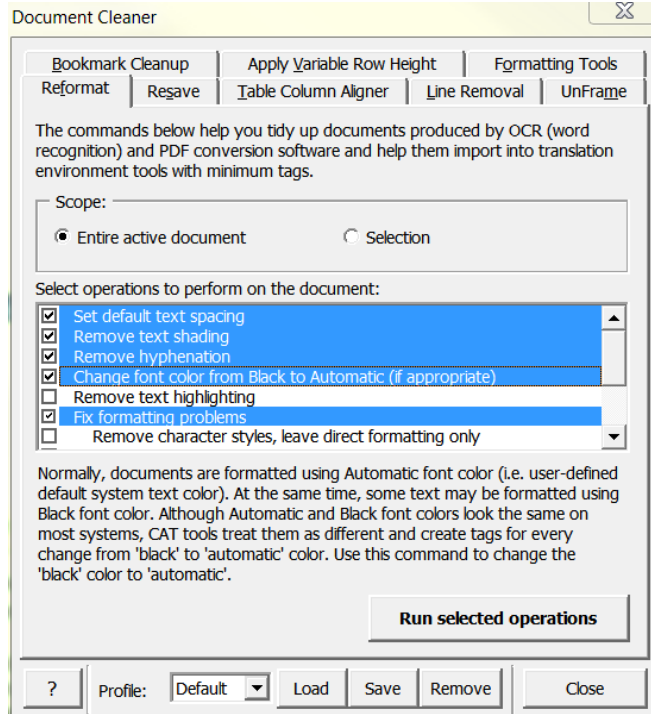
files, you can choose to use [inline tags](#) which can be reordered and omitted, as necessary.

- [Format](#) menu, keyboard shortcuts, toolbar buttons
- Tags can be edited
- WordFast
  - Edit > [Tags](#) menu, keyboard shortcuts, toolbar buttons or manually typed
  - Tags can be edited

#### 4.5 Tag verification

A quality assurance (QA) feature to verify that tags have been used correctly (i.e. based on the verification settings)

- Trados Studio
  - Each file type has its own [tag check settings](#) which can be accessed either via Project Settings or Options (see Figure 5)
- memoQ
  - [QA settings](#) > Inline tags (see Figure 6)
- WordFast
  - Transcheck feature can check for missing tags but there are no customizable settings for this



**Figure 1.** Document Cleaner – settings for the Reformat function.

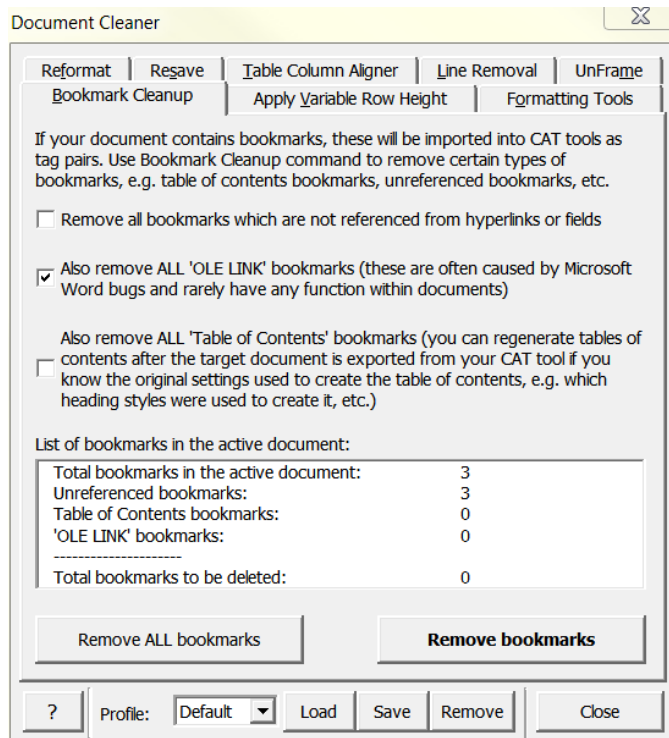


Figure 2. Document Cleaner – settings for the Bookmark Cleanup function.

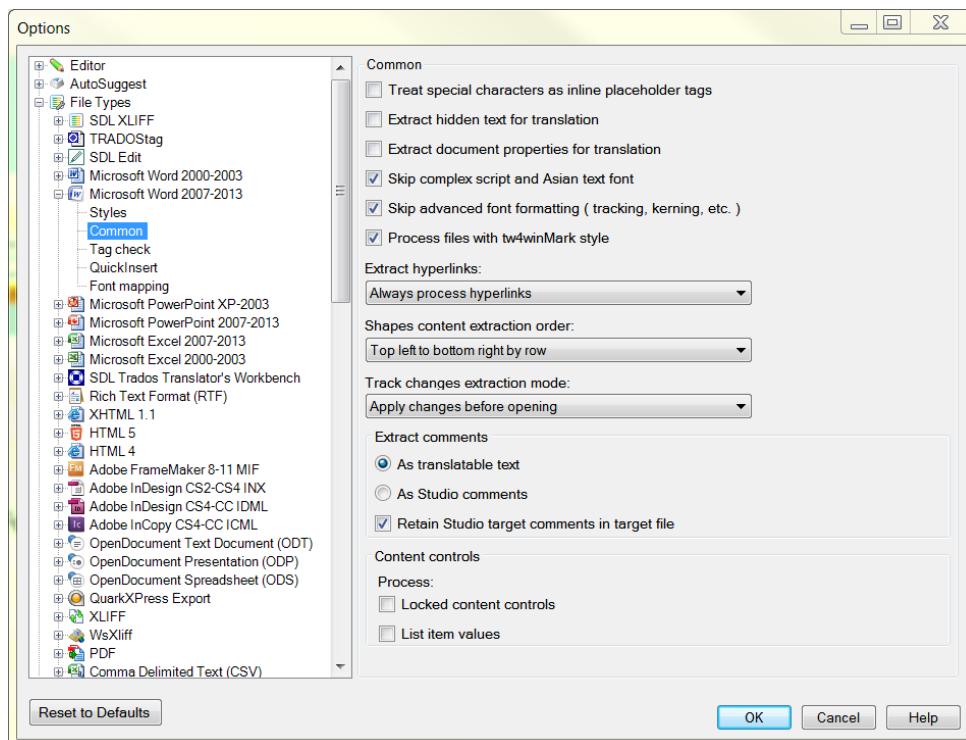
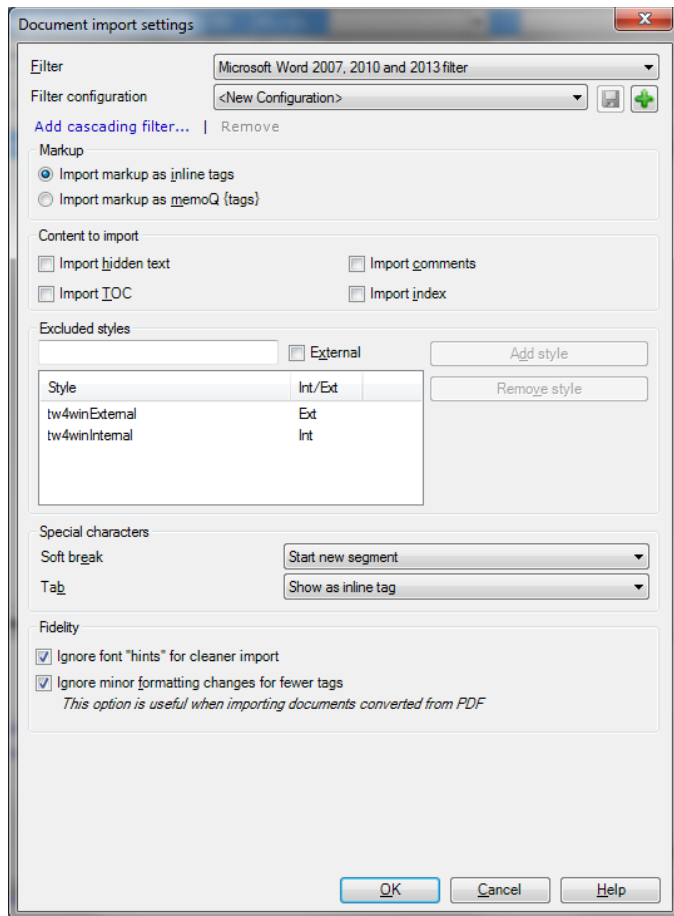
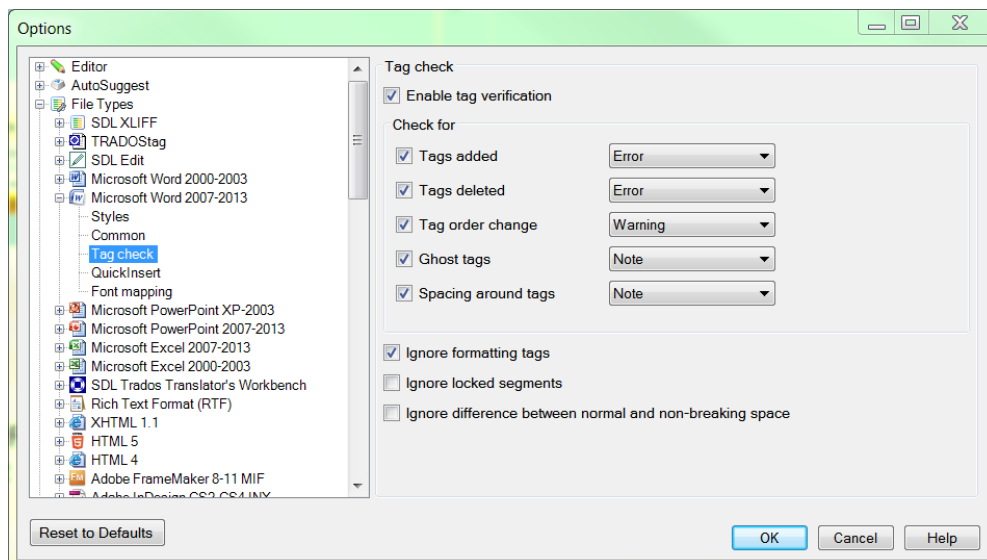


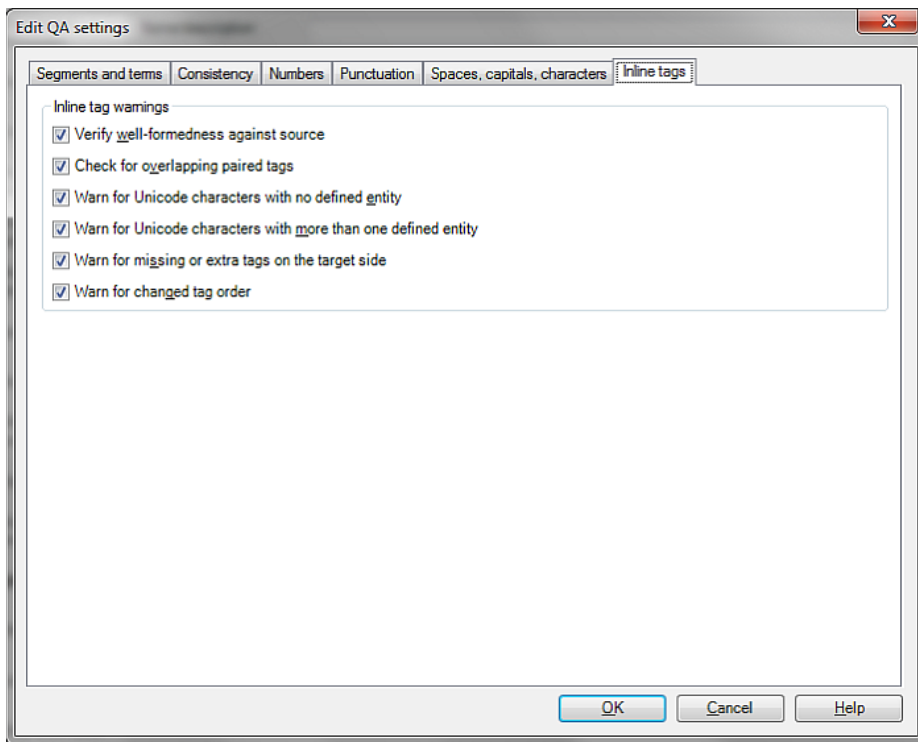
Figure 3. File type settings for DOCX files in Trados Studio 2014.



**Figure 4.** Document import settings for DOCX files in memoQ 2013.



**Figure 5.** Tag verification settings for DOCX files in Trados Studio 2014.



**Figure 6.** QA settings for inline tags in memoQ 2013.