

# XLIFF:doc

## The XLIFF Representation Guide for Documents

<b>Introduction to XLIFF:doc</b>	<b>4</b>
Why Use XLIFF:doc?	4
Getting Started	5
Conventions	5
<b>Using XLIFF:doc</b>	<b>6</b>
<b>Project Meta Data</b>	<b>7</b>
<b>Organizing Source Files into XLIFF:doc Files</b>	<b>8</b>
Multiple Source Files vs Single Source File	8
One Target Language Per XLIFF	8
<b>Structure (Format), Content (Text), and Previews</b>	<b>9</b>
Overview	9
Skeleton (Structure)	9
Content (Text)	9
Whitespace Between Translation Units	9
HTML Previews	10
XSL Previews	11
<b>Joining and Splitting Translation Units</b>	<b>12</b>
Limiting Joining and Splitting at the File Level	12
Limiting Joining at the Translation Unit Level	12
<b>Describing Translation State</b>	<b>13</b>
Translation Status	13
Recording Match Quality After a Match has been Applied	14
Last Edited Information	14
Usage Example: Translation Attributes and Workflows	14
<b>Inline Formatting</b>	<b>16</b>
Interpreting Tag Content for Users	16
Providing Text Representations of Standalone Tags	16
Providing HTML Representations of Tags	17
<b>Notes</b>	<b>18</b>
<b>Translation Matches</b>	<b>19</b>
Rules for TM Matches in <alt-trans>	19
Describing Match Quality	19

## Interoperability

Providing Context for Matches	20
Describing Attributes of the Match	21
Usage Example	21

### Terminology Matches 23

Term Embedding Mechanism	23
Embedding Term Matches	23
UTX as Supported by XLIFF:doc	24
Reading and Displaying Embedded Term Matches	25

### QA Information 26

Usage Example 1	26
Usage Example 2	27

### Revision History 28

Unit Histories	28
----------------	----

## Appendix A: Diagram of XLIFF:doc Elements 29

## Appendix B: Reference Guide to XLIFF:doc 30

<alt-trans>	30
<body>	31
<dx:attribute>	31
<dx:attributes>	31
<dx:content>	32
<dx:context>	32
<dx:context-source-after>	32
<dx:context-source-before>	32
<dx:context-target-after>	33
<dx:context-target-before>	33
<dx:external-xsl>	33
<dx:internal-preview>	33
<dx:qa-hit>	34
<dx:note>	35
<dx:qa-hits>	36
<dx:utx-comment>	36
<dx:utx-glossary>	36
<dx:utx-src>	37
<dx:utx-term>	37
<dx:utx-tgt>	38

## Interoperability

<external-file>	38
<file>	38
<g>	40
<group>	41
<header>	41
<mrk>	42
<skl>	42
<source>	42
<target>	42
<tool>	43
<trans-unit>	43
<x>	45
<xliff>	46

<b>Appendix C: Version History</b>	<b>48</b>
------------------------------------	-----------

# Introduction to XLIFF:doc

XLIFF:doc is a reference guide to representing documents in XLIFF files. The primary goal of XLIFF:doc is to provide a mechanism to allow lossless exchange of XLIFF data between tools in the localization workflow.

Interoperability is achieved by:

- Using the shared namespace extension dx: with defined attributes and attribute values.
- Using only a subset of the XLIFF 1.2 specification, with pre-defined attribute values.

☼ *Note: XLIFF:doc is not intended for use as a permanent or semi-permanent bilingual project file format. Its design is focused solely on being an efficient, temporary mechanism for exchanging data between two systems. If your goal is to exchange data between 3 or more tools, the appropriate use of XLIFF:doc in that scenario is to have each tool create a new XLIFF:doc and pass it to the next tool in line, or to use one tool as the hub in a hub-and-spoke wheel arrangement.*

XLIFF:doc is designed to support the following file types:

- XML document file formats (DITA, structured Framemaker, etc.)
- Office documents (.doc, .docx, .rtf, .otd, .xls, etc.)
- FrameMaker
- InDesign
- HTML and XHTML

XLIFF:doc may not be ideal for every document type or localization challenge. For software string formats, the XLIFF:software representation guide is recommended. For web and interactive formats (Flash, HTML5, etc.), the XLIFF:web representation guide is recommended.

☼ *Note: the above paragraph contains forward looking statements. As of this time, XLIFF:software and XLIFF:web do not exist.*

## Why Use XLIFF:doc?

XLIFF:doc is not the only way to write a bilingual localization file. It is not the only way to write XLIFFs. So why should you support it in your tool?

- Interoperability: Any tool that follows the XLIFF:doc representation guide can read and write XLIFF:doc files from your tool, with no data loss
- Less is more: XLIFF:doc only uses a fraction of the XLIFF 1.2 specification, so you can have a robust XLIFF solution without needing to support every last element and attribute of the XLIFF format.
- Tailor-made for documentation: XLIFF:doc was designed specifically to handle the translation of documents. The elements, attributes, and attribute values in the XLIFF:doc namespace ("dx:") bridge the gap between the promise of the XLIFF format, and the reality of needing to store more information than is provided for by the XLIFF 1.2 specification.
- Embedded terminology: XLIFF:doc provides an easy way for tools that are not terminology-aware, and are not hooked up to terminology databases, to show the linguist relevant terminology hits.
- Create previews from any tool: Any XLIFF:doc-aware tool can generate formatted previews that approximate the final file format, using either source or translation text.

## Interoperability

- Fully-documented: You do not need to reverse engineer anything in order to start using XLIFF:doc. All elements, attributes, and even attribute values are fully documented, both for the subset of the XLIFF spec. used, and for the XLIFF:doc custom namespace extension.

## Getting Started

It is not necessary to study the XLIFF 1.2 specification in detail before reading this document. However, readers should be comfortable with the relevant portions of the XLIFF 1.2 specification before developing any solutions based on XLIFF:doc.

The Using XLIFF:doc chapter is meant to be a guide to creating and reading XLIFFs. The next two chapters contain in-depth reference on the individual elements and attributes that make up the XLIFF:doc representation guide.

## Conventions

Throughout this document, "Creation Tool" refers to a tool that creates an XLIFF:doc file from one or more source files. This would include segmentation, it may or may not include TM matching, TB matching, pre-translation, etc.

"Processing Tool" refers to any tool that reads information from an already-created XLIFF:doc file. Tools that modify XLIFF:doc files are Processing Tools, in the scope of this document.

Names of XLIFF and XLIFF:doc elements are formatted as follows: `<trans-unit>`.

Names of XLIFF and XLIFF:doc attributes are formatted as follows: `dx:qa-hit`.

## Using XLIFF:doc

### Project Meta Data

XLIFF:doc does not currently have a dedicated mechanism to store meta data about the project an XLIFF:doc file belongs to. If you need to convey project information from tool to tool, you may wish to look into using the Translation Interoperability Protocol Package (TIPP) format. The TIPP format has well-defined mechanisms for meta data about the project, and allows for an XLIFF:doc to be embedded into the TIPP file.



# Organizing Source Files into XLIFF:doc Files

## Multiple Source Files vs Single Source File

You can bundle multiple source files into a single XLIFF:doc, or you can use a model of one XLIFF:doc per source file.

## One Target Language Per XLIFF

Tools processing XLIFF:doc files are expecting exactly one target language per XLIFF:doc. You must not use multiple `<file>` elements to include multiple versions of the source file (one for each target language in the project). If you have multiple target languages, you must create a different XLIFF:doc for each one.

## Structure (Format), Content (Text), and Previews

### Overview

In most interpretations of XLIFF, non-inline formatting data (structural data) is separated from the text data, and put into a separate file or element called a skeleton. Text data is separated from structural data, and placed into `<trans-unit>` elements. Once the file is translated, the data in the `<target>` element of each `<trans-unit>` is merged back into the skeleton to generate a translated file in the original file format. In an ideal world, every tool that processed an XLIFF file would have the capability to generate a formatted target language file. However, due to the complexity and variety of file formats that need translation, it is not easy to provide a skeleton file that can be used universally. Recognizing this, XLIFF:doc adds a preview element into which the Creation Tool is required to place enough information for any compliant tool to show the user a visual preview.

### Skeleton (Structure)

In XLIFF:doc, skeleton files are stored outside of the XLIFF itself, in .skl files. Each XLIFF file will have a corresponding .skl file. This skeleton file can be placed alongside the XLIFF in the translation package, or it can be left on the server or in another location. This choice is up to the user or tool that creates the XLIFF. XLIFF:doc does not require the skeleton file to be available to every tool or user of the XLIFF. XLIFF:doc does not specify how a skeleton file should be created, stored, or structured.

### Content (Text)

How exactly a Creation Tool extracts text from a source file and organizes it into `<trans-unit>` elements is not, in general, specified by XLIFF:doc. There is one exception, however: subflows. XLIFF 1.2 allows for subflows, to handle situations like a sentence containing an HTML link which has a title that must be translated in addition to the text of the sentence. For example:

January 24 is `<a href="http://www.allyourpeanutbutterarebelongtous.com/index.html" title="Somebody set up us the bomb">`National Peanut Butter Day`</a>` in the United States.

XLIFF:doc does not support the subflow `<sub>` element. To handle text that would be considered a subflow, extract it into a separate translation unit. For example, the above sentence might be rendered into 2 translation units as follows:

```
<group id="6">
  <trans-unit id="9">
    <source>January 24 is <g id="18" ctype="link" dx:orig-markup-open="&lt;a
href=&quot;http://www.allyourpeanutbutterarebelongtous.com/index.html&quot; title=&quot;Somebody
set up us the bomb&quot;&gt;" dx:orig-markup-close="&lt;/a&gt;">National Peanut Butter Day</g> in
the United States.</source>
    <target state="new"></target>
  </trans-unit>
</group>
<group id="7">
  <trans-unit id="10">
    <source>Somebody set up us the bomb</source>
    <target state="new"></target>
  </trans-unit>
</group>
```

☼ Note: XLIFF:doc does not specify whether the subflow part comes before the TU containing the reference to the subflow. In other words, the contents of TU 9 and TU 10 could have been swapped in the example above.

### Whitespace Between Translation Units

In order to allow processing tools to merge TUs correctly, and to generate accurate previews, XLIFF:doc requires the Creation Tool to remove any end-of-segment whitespace between 2 merge-able TUs. It is the responsibility of the

## Interoperability

Processing Tools to insert the appropriate whitespace between segments when generating a preview or target document.

☀ Example. Assume the following piece of source text:

[start of paragraph]This is a hat. The hat is blue. I like the hat.[end of paragraph]

After segmentation and placement in an XLIFF:doc, the whitespace from the end of the first two sentences has been removed, and does not make it into the XLIFF:doc at all, as follows. Note that the last segment for the last sentence in the paragraph did not have any whitespace at the end.

```
<source>This is a hat.</source>
<source>The hat is blue.</source>
<source>I like the hat.</source>
```

For a French translation, the targets may end up like this:

```
<target>C'est un chapeau.</target>
<target>Le chapeau est de couleur bleue.</target>
<target>J'aime le chapeau.</target>
```

A preview generating by a Processing Tool, or a final target document generated by the Creation Tool would add appropriate spaces back in:

[start of paragraph]C'est un chapeau. Le chapeau est de couleur bleue. J'aime le chapeau.[end of paragraph]

For Japanese, a language with no spaces between sentences, the `<target>` elements may end up as below:

```
<target>これは帽子です。</target>
<target>この帽子は青です。</target>
<target>この帽子が好きです。</target>
```

A preview generating by a Processing Tool, or a final target document generated by the Creation Tool would abstain from adding spaces to the output document, since they are not appropriate Japanese text (note that Japanese full stop characters typically have some whitespace built into them):

[start of paragraph]これは帽子です。この帽子は青です。この帽子が好きです。[end of paragraph]

## HTML Previews

While XLIFF:doc does not guarantee that any compliant tool can generate a target from every XLIFF:doc, it does guarantee that any compliant tool can generate an HTML view of the original document (with source and/or target text showing). The fidelity of the preview could range from a simple table view, to an approximation of a document, with headers, bold text, images, etc., to a high-fidelity preview that is functionally and visually identical to the original file format. XLIFF:doc does not put any requirements around how "similar" a preview must be to the original file format. Fidelity will be determined by the complexity of the original file format, and the effort put into creating the preview by the Creation Tool. The preview is embedded into the XLIFF file.

The following items give an overview of the preview:

1. The preview will be constructed in HTML format. There are no requirements to use a specific version of HTML. However, cross-platform, cross-browser code is highly recommended (YAML, etc.).
2. The preview can reference images, but only if those images are either external (via URL) or relative to images included in a TIP package (if the XLIFF:doc is part of a TIPP file).
3. The HTML for the preview will be escaped, as in this example:

```
&lt;t;html&gt;
&lt;t;head&gt;
```

## Interoperability

```
<title>Translation Preview</title>
</head>
```

4. A `<dx:content>` element will be placed in the preview at any location translatable text has been extracted out into `<trans-unit>` elements. The TU ID of the `<trans-unit>` in question will be specified in the *id* attribute of the `<dx:content>` element, as follows, as in this example for TU ID "2":

```
<div id="siteSub"><dx:content id="2"/></div>
```

### Generating an HTML Preview from an XLIFF

The general outline for creating an HTML preview might resemble the following:

1. Obtain contents of the `<dx:internal-preview>` element and unescape it.
2. Replace `<dx:content>` elements with the contents of the `<target>` element. For inline tags, use the *dx:equiv-markup* and *equiv-text* attributes as appropriate. Do not use original markup unless the original file format was HTML. If no equivalent information is available for a placeholder, consider substituting a generic character or string (eg, "{tag}" for standalone placeholders, and underlining for paired tags. In that way, at least some information will be provided on the position of placeholders.
3. Display the resulting HTML to the user

The details of implementation are not specified by XLIFF:doc. You could use XSLT to accomplish it, or string-based code.

### XSL Previews

In addition to HTML previews, XLIFF:doc provides a second, optional preview type based on XSLT. Creator Tools are not required to support this preview mechanism. If a Creator Tool does provide for an XSL preview, it must, however, supply the XSLT needed to generate the preview, and any Processing Tool that opens the XLIFF:doc must be able to generate a preview for the user using that XSL.

The following items give an overview of the XSL preview:

1. The Creator Tool embeds a link to an XSLT file relative to the XLIFF:doc file. This is done using the `<dx:xsl-external-file>` element.
2. The Creator Tool populates the *dx:xsl-hook* attribute in each `<trans-unit>`.
4. To generate a preview, a Processing Tool obtains the XSLT and applies it to the XLIFF:doc file itself, then displays the resulting HTML to the user.

### Joining and Splitting Translation Units

XLIFF:doc supports the joining and splitting of translation units (segments). Creation Tools can decide whether or not to allow it for specific packages, source files, and individual translation units. Processing Tools must support joining and splitting, but must also respect the limitations set for a file by the Creation Tool. It is the responsibility of the Processing Tools to ensure that inline formatting tags (<g> and <x> placeholders) and inter-segment white spaces are handled appropriately during split and join operations.

#### Limiting Joining and Splitting at the File Level

A Creation Tool can allow or disallow splitting and joining for an entire source file by setting the *dx:allow-tu-join* and *dx:allow-tu-split* attributes of the <file> element. Both attributes default to "yes", so they only need to be set if you do not want to allow joining and/or splitting. Processing Tools may not adjust these values, and must enable/disable joining and splitting accordingly to be compliant to XLIFF:doc.

#### Limiting Joining at the Translation Unit Level

Even if joining of TUs is permitted in general for a file, there may be specific reasons why certain TUs should not be joined. Typically, this would be because there is structural markup between the TUs that would cause the document to lose stability if the TUs were joined. In XLIFF:doc, the <group> element is used only to prevent specific TUs from being joined. The <body> element needs at least 1 <group> element, and all TUs have to be part of a <group>. A TU that is not in the same group as another TU cannot be joined with that TU.

- ☼ Example 1: If you had a purely text file (no markup), you would likely end up with a single <group> for the entire XLIFF. That group would contain all the TUs in the file, and any TU could be joined with the one before or after it.
- ☼ Example 2: If you had a simple XML document with 3 <para> elements, you might have three <group> elements in your XLIFF, one for each <para> element in the original document. That would allow the sentences within the para to be joined with each other, but joining would be prevented across the <para> element.
- ☼ *Note: Add something about where store restype, story ids, etc as used by Rainbow for InDesign docs. Will apply these to the TU object, and not add another grouping level.*

Note: if splitting is permitted for the file as a whole, it is permitted for every TU in the document, and cannot be overridden on a per-TU basis.

## Describing Translation State

In the XLIFF 1.2 specification, each translation unit in an XLIFF file can have `<alt-trans>` translation matches, each of which has a preset match-rate described for it. Each project translation unit can also be assigned a "state", such as "new", "translated", etc. There is not, however, a built-in mechanism to describe the quality of the translation unit after a machine translation has been applied, or after an embedded TM match or a live TM lookup match has been applied. If you save an XLIFF after applying a match, no information is saved about where the translation came from, how close a match it was to the source, etc.

XLIFF:doc addresses these shortcomings in two ways: specifying exactly which status labels can be applied to a translation unit, and by adding new attributes to the `<trans-unit>` to enable it to maintain its own state better. Some information is stored on the `<trans-unit>` level, and some is stored on the `<target>` level.

## Translation Status

XLIFF:doc uses the `state` attribute of the `<target>` to record the state of a translation. The XLIFF 1.2 specification includes 10 pre-defined states, and allows for unlimited user-defined states. This has meant, in practice, that very few tools and organizations use the same states. Or, if they do use the same states, they often don't use them the same way (to mean the same thing). To help foster interoperability between tools, XLIFF:doc specifies a much smaller set of states: "new", "translated", "final", "signed-off", and "needs-review-translation". Use of other states is not permitted within XLIFF:doc. How individual organizations interpret these statuses is outside the purview of XLIFF:doc, but one possible interpretation is as follows:

Description	XLIFF:doc state	UI Representation
New translations in a new XLIFF	new	New
A complete translation is present in the target	translated	Translated
The translation has been validated for accuracy, context, terminology, etc., according to the procedures and policies of the organization	final	Proofed
The final authority on the target language (a client's subject matter expert, for example) has reviewed the translation and approved it.	signed-off	Validated
An expert or other authority has reviewed the translation and rejected the proposed translation. This TU must be reviewed and modified as necessary.	needs-review-translation	Rejected

## Addition Notes on Status Usage

- XLIFF:doc does not require any particular usage of the statuses provided. The examples provided in this document are just that: examples. However, statuses are important to passing projects smoothly between different users and groups, and so, to encourage more interoperability, XLIFF:doc limits the available statuses to the ones listed above.
- Depending on process, translations from a validated TM source that are placed during pre-translation by the Creation Tool may be set to "final" or "signed-off" from the start.
- Depending on process, setting the status of "translated" may play a role similar to "confirming" a translation in tools that support an on/off state for confirmation.
- The reviewer does not necessarily have to add a note when they mark a translation as Rejected (needs-review-translation), but in an ideal world, perhaps they would. This is not something XLIFF:doc specifies.
- The usage of the "final" and "signed-off" states may vary from situation to situation and organization to organization.
- For an organization with a simple Translation-Edit-Proof (TEP) process, The "Translate" step may end with all TUs marked "translated", the "Edit" step with all TUs marked "final", and the Proof step with all TUs marked as "signed-off".

## Interoperability

- For an organization doing TEP followed by a client review, the Translation stage may end at "new", the Edit phase at "translated", the Proof step at "final", and the client's review phase at either "needs-review-translation" or "signed-off". Once all rejected TUs had been successfully reviewed and review comments were incorporated, all TUs might then be marked "signed-off".
- What "final" means to an organization will vary, but the basic idea should be the same: this translation is certified as accurate and correct for the context it is in, according to the procedures of the organization handling the translation.

## Recording Match Quality After a Match has been Applied

If a translation has been applied to the TU, and the user has not subsequently edited the `<target>`, then information about that match can be recorded against the `<trans-unit>`. This allows users to identify how a translation was populated, whether by human editing, or from a live TM, live machine translation engine, embedded XLIFF match, etc. Use the `dx:match-quality`, `dx:glorious-match`, `dx:origin`, `dx:origin-shorttext`, and `dx:match-penalty` to imprint the TU with the status of the match applied to it.

☼ *Note: if a translator edits a translation, the match-related information must be changed, as it no longer accurately reflects the match that was applied. `dx:glorious-match`, `dx:match-quality`, and `dx:match-penalty` must be removed. `dx:origin` must be set to "manual" and `dx:origin-shorttext` to "Manual Translation".*

## Last Edited Information

Use `dx:modified-by` and `dx:modified-at` to show when the TU was last edited, and who edited it. When a Creation Tool places translations ("pre-translation") in an XLIFF:doc, it must either specify the user who last edited the translation in the TM, or a general identifier (eg, "Pre-Translation", "Acme TM Server", etc.).

## Usage Example: Translation Attributes and Workflows

XLIFF:doc does not define translation workflow. This example is provided merely as an example of how actions in real world workflows might be reflected in XLIFF:doc.

☼ *Note: The UI for a given tool can choose to represent the permitted translations states in whatever manner they want. In this example, "needs-review-translation" is shown as "Rejected", and "signed-off" is shown as "Validated".*

Step #	Workflow Step	XLIFF:doc Representation
1	TM server pre-translates file, finding exact, in-context matches for 25% of the TUs. The matches are coming out of a TM where all translations have been reviewed and validated.	<code>&lt;target&gt;</code> element is populated with translation from TM.  <code>target/state</code> is set to "translated"  <code>trans-unit/dx:match-quality</code> is set to "100"  <code>trans-unit/dx:glorious-match</code> is set to "yes"
2	TM server pre-translates file, finding exact matches for 50% of the TUs.	<code>&lt;target&gt;</code> element is populated with translation from TM.  <code>target/state</code> is set to "translated"  <code>trans-unit/dx:match-quality</code> is set to "100"

## Interoperability

Step #	Workflow Step	XLIFF:doc Representation
3	TM server finds no matches or only fuzzy matches for the remaining 25% of the TUs.	<p>&lt;target&gt; element is left empty.</p> <p>target/state is set to "new"</p> <p>&lt;alt-trans&gt; elements are added with fuzzy matches</p>
4	Translator reviews the matches listed as "Translated", and finds some that need to be modified to match context, subject matter, etc. Translator makes appropriate changes. Since these are already set to "Translated", Translator doesn't have to change their status when she edits them.	<p>target/state stays at "translated".</p> <p>trans-unit/dx:match-quality is removed (it is now a manual translation).</p>
5	Translator sees that 25% of the translations are in-context exact matches already marked as "Translated", and so skips past these translations.	n/a
6	Translator filters to "New" translations, then for each one, manually translates the source, or transfers fuzzy matches to the target, and modifies as needed. Translator sets status to "Translated"	<p>target/state is set to "translated"</p> <p>trans-unit/dx:match-quality is removed (it is now a manual translation).</p>
7	Editor receives file and confirms that all TUs are set to "Translated"	[all TUs have already been set to target/state="translated"]
8	For each translation he disagrees with, the Editor changes the translation as appropriate. Since the Editor is not planning on sending these back to the first linguist for approval, the Editor sets the status of these translations as "Proofed".	<p>target/state is set to "final"</p> <p>If the the translation edited was previously a 100%, ICE, or fuzzy match, the trans-unit/dx:match-quality is removed (it is now a manual translation).</p>
9	For each translation he agrees with, the Editor changes status to "Proofed".	target/state is set to "final"
10	Reviewer receives file and confirms that all TUs are set to "Proofed"	[all TUs have already been set to target/state="final"]
11	For each translation he disagrees with, the Reviewer adds a comment, and changes status to "Rejected".	target/state is set to "needs-review-translation" and a <dx:note> element is added to the <trans-unit> element.
12	For each translation he agrees with, the Reviewer changes status to "Validated".	target/state is set to "signed-off"
13	File is returned to Translator, who checks all the TUs marked as Rejected. If she agrees with them all, she incorporates the suggestions found in the comments, and marks the TUs as "Validated". Any comment suggestions that she is unable to incorporate, she leaves the TU status as "Rejected", but adds a note to explain why the change would not be appropriate. Steps 11-13 then repeat until all differences have been resolved and all TUs have been marked as "Validated".	target/state is set to "signed-off" or a <dx:note> element is added to the <trans-unit> element.



## Inline Formatting

In order to simplify the handling of inline formatting, all inline formatting in XLIFF:doc is handled via the `<g>` and `<x>` tags. The other elements provided by the full XLIFF 1.2 specification are not supported by XLIFF:doc. The original formatting must be included in the XLIFF. In addition, the *equiv-text* attribute must be specified when working with `<x>` tags.

## Interpreting Tag Content for Users

XLIFF:doc can be used to represent content from hundreds of content creation systems, including not just commercial applications like MS Word, Adobe InDesign, Adobe FrameMaker, but also HTML and XML-based CMS systems. Obviously, inline formatting can vary widely from system to system. To help linguists correctly place tags when working in XLIFF:doc, it is important that XLIFF:doc Creation Tools provide as much information as possible about the inline formatting captured in `<g>` and `<x>` elements. In XLIFF:doc, this is accomplished by using the *ctype* attribute. Use the list of pre-defined values for *ctype* provided in the Reference Section.

☀ *Note: Although XLIFF:doc does not specify exactly how to map original formatting code to the predefined list of *ctype* values, the more effort a Creation Tool makes to map detailed formatting codes to easy-to-understand *ctype* values, the better able Processing Tools will be able to represent the document to users, and the better able the users will be to place the tags correctly. As an example, the following tags might all be described as "bold" in XLIFF:doc: RTF {b}, MS Word <w:b>, HTML <b>, HTML <strong>, Custom XML <Bold>, Custom XML <TotallyCritical>.*

## Providing Text Representations of Standalone Tags

In order to provide linguists with the information they need to evaluate a segment in XLIFF:doc, use the *equiv-text* attribute with your `<x>` tags. The idea behind *equiv-text* is to provide at least a rough approximation of how placeholders will be resolved in the final system. For example, if the `<x>` tag encapsulates an XML entity that a CMS style sheet will resolve to a unit of measurement in the final document, you might populate the *equiv-text* value with that unit of measurement. If the `<x>` tag represents a line-break, you might choose to insert a line-break symbol (or a space, etc.). If the `<x>` tag represents a non-breaking space in the source system, you might use a space character in *equiv-text*.

### equiv-text Mapping Examples

Encapsulated Content	Meaning in Source System	equiv-text Value
&sys-volts;	Unit of Measurement: Volt	V
\n	Line break	¶
&sys-product-name;	Name of product	ProductName
<company-name type="non-local"/>	Name of company	XYZ Corporation, Inc.
<w:page-num/>	Current page number	#

☀ *Note: XLIFF:doc does not specify how you populate the *equiv-text* attribute, only that you do it. The above examples are not meant to be guidelines.*

### equiv-text Readability Examples

Simplified Tags	No equiv-text Provided	With equiv-text
Ensure that the defibrillation lead impedance is greater than 20{1}.	Ensure that the defibrillation lead impedance is greater than 20.	Ensure that the defibrillation lead impedance is greater than 20 Ω.
69{1} x 51{2} x 15{3}	69 x 51 x 15	69 mm x 51 mm x 15 mm

## Interoperability

Simplified Tags	No equiv-text Provided	With equiv-text
The Model {1} {2} device and the pacing leads and defibrillation leads constitute the implantable portion of the device system.	The Model device and the pacing leads and defibrillation leads constitute the implantable portion of the device system.	The Model Acme SuperDefibrillator device and the pacing leads and defibrillation leads constitute the implantable portion of the device system.
For a list of the features that are enabled at shipping, see the "Shipped" column of the tables in {1}.	For a list of the features that are enabled at shipping, see the "Shipped" column of the tables in .	For a list of the features that are enabled at shipping, see the "Shipped" column of the tables in {XREF}.
(02)-123-4567{1}Fax (02)-123-5678	(02)-123-4567Fax (02)-123-5678	(02)-123-4567 Fax (02)-123-5678

## Providing HTML Representations of Tags

To give Processing Tools the ability to show formatted XLIFF:doc content to users, XLIFF:doc uses the *dx:equiv-markup*, *dx:equiv-markup-open*, and *dx:equiv-markup-close* attributes. Populate these attributes with HTML code that approximates the formatting of the original document format.

☼ *Note: if the original format is HTML, and the HTML formatting can be represented safely, it is not necessary to include *dx:equiv-markup* attributes. If not present, the *dx:orig-markup* will be used. See the section on *ctype* in the reference section for more information.*

See the table below for examples:

Original Document Formatting	XLIFF:doc Example
HTML: This is <b>important</b> .	This is <code>&lt;g id="1" ctype="bold" dx:orig-markup-open="&amp;lt;em&amp;gt;" dx:orig-markup-close="&amp;lt;/em&amp;gt;"&gt;important&lt;/g&gt;</code> . Might be rendered as: This is <b>important</b> .
XML: This is <b>important</b> .	This is <code>&lt;g id="1" ctype="bold" dx:orig-markup-open="&amp;lt;emphasis&amp;gt;" dx:orig-markup-close="&amp;lt;/emphasis&amp;gt;" dx:equiv-markup-open="&amp;lt;em&amp;gt;" dx:equiv-markup-close="&amp;lt;/em&amp;gt;"&gt;important&lt;/g&gt;</code> . Might be rendered as: This is <b>important</b> .
Docx: <code>&lt;w:r&gt;&lt;w:t&gt;This is&lt;/w:t&gt;&lt;/w:r&gt;&lt;w:r&gt;&lt;w:rpr&gt;&lt;w:b /&gt;&lt;/w:rpr&gt;&lt;w:t&gt;Important&lt;/w:t&gt;&lt;/w:r&gt;</code>	This is <code>&lt;g id="1" ctype="x-other" dx:orig-markup-open="&amp;lt;&amp;lt;w:r&amp;gt;&amp;lt;w:t&amp;gt;&amp;gt;" dx:orig-markup-close="&amp;lt;/w:t&amp;gt;&amp;lt;/w:r&amp;gt;" dx:equiv-markup-open="&amp;lt;" dx:equiv-markup-close="&amp;gt;"&gt;&lt;g id="2" ctype="bold" dx:orig-markup-open="&amp;lt;w:r&amp;gt;&amp;lt;w:rpr&amp;gt;&amp;lt;w:b /&amp;gt;&amp;lt;/w:rpr&amp;gt;&amp;lt;w:t&amp;gt;" dx:orig-markup-close="&amp;lt;/w:t&amp;gt;&amp;lt;/w:r&amp;gt;" dx:equiv-markup-open="&amp;lt;b&amp;gt;" dx:equiv-markup-close="&amp;lt;/b&amp;gt;"&gt;important&lt;/g&gt;</code> Might be rendered as: This is <b>important</b>

### Notes

In XLIFF:doc, users can apply notes to two distinct parts of the XLIFF:

- a specific translation unit (either a current `<trans-unit>` or a leveraged `<alt-trans>`)
- the entire file

 *Note: Notes that refer to the file as a whole are entered against the `<header>` element.*

Notes are stored with the name or description of the user/system that inserted or last modified the note, and the last date/time associated with the note. XLIFF:doc uses the `<dx:note>` element to store the actual note text, and the `dx:modified-by` attribute to define who created or last edited the note. The `dx:modified-at` attribute is used to define when the note was created or last edited.

Notes that refer to a specific TU can optionally refer to either the source or target. Unless specified otherwise, notes refer to the TU as a whole.

XLIFF:doc does not support notes tied to other XLIFF elements.

### Translation Matches

When translating an XLIFF:doc file, translation memory matches can come from two sources:

1. A "live" TM: a TM available from a networked server, or a file-based TM the translators has on their desktop.
2. Embedded matches: TM matches pre-placed in the XLIFF:doc by a CAT tool.

☼ *Note: The use of live TMs to find translation matches is outside of the scope of XLIFF:doc, and is defined by the Processing Tool.*

Under XLIFF:doc, TM matches are embedded per-TU, using the `<alt-trans>` element. Any embedded matches will be available to any translator editing the XLIFF:doc using a compatible tool, regardless of whether or not they have access to a TM server or database file.

### Rules for TM Matches in `<alt-trans>`

XLIFF:doc establishes several rules for embedded TM matches, and allows a Creation Tool to set other rules. These are the fixed rules which must be followed by any XLIFF:doc compliant tool:

- Only the Creation Tool can set the `dx:allow-tu-match-deletion` and `dx:allow-tu-match-addition` attributes
- Processing Tools must retain any TM matches embedded in an XLIFF:doc by the Creation Tool unless the `dx:allow-tu-match-deletion` attribute is set to "yes". If an XLIFF:doc file has 500 TM matches in it when a tool opens it, the tool must write it out with those same 500 matches.
- Processing Tools must not add TM matches to the XLIFF:doc, unless the `dx:allow-tu-match-addition` attribute is set to "yes". Processing Tools can always show users matches from other sources, but it cannot pass these matches along to subsequent tools unless permitted to do so by the Creation Tool.
- Processing Tools must display all embedded TM matches to users when the user requests them.
- Processing Tools must maintain the relationship between an embedded TM match, and the TU that it was originally assigned to. If a Processing Tool reads an XLIFF:doc into some other bilingual format native to the tool, and in doing so adds the embedded TM matches to a database, then the tool must retain the link between individual TM entries and the XLIFF TUs the Creation Tool embedded them into. The tool must display the formerly embedded matches relevant to the TUs they were originally embedded in, whenever the user requests to see matches for the TU. See the next rule.
- Processing Tools must display embedded TM matches associated with a particular TU, regardless of the user's live TM settings. For example, if a user has chosen to only display TM matches of 85% or higher, the tool must still display a 50% match to the user, if a 50% match was embedded for the TU in question. The reason for this is to honor the settings of the original TM server (Creation Tool) that placed the matches. However, a tool may have a user-controlled switch that toggles the display of embedded matches on and off.

☼ *Note: these limits apply only to TM matches in alt-trans. They do not apply to alt-trans used for revision history purposes.*

### Describing Match Quality

The `match-quality` attribute is used to indicate how similar a translation candidate's source text is to the XLIFF TU's source text. Use values 0-100 (decimal numbers are acceptable).

☼ *Note: match-quality and dx:match-quality refer solely to the similarity of the source text to the match's source text.*

To indicate an in-context, exact match, use a `dx:match-quality` of 100, and set `dx:glorious-match` to "yes".

## Interoperability

If the Creation Tool assigns an artificial penalty to translation matches, the amount of that penalty can be recorded in the *dx:match-penalty* attribute. Do not specify textual difference penalties here: this value is used to describe penalties associated with the match's origin or attributes.

- ☀ Example 1: An exact match is found in the TM, and the context matches. The match is from the same product line and business unit. The *match-quality* is set to 100%, *dx:glorious-match* is set to "yes", and *dx:match-penalty* is set to 0%. The CAT tool displays this match to the user as a 100% match.
- ☀ Example 2: An exact match is found in the TM, but the match is from a different product line and context. Additionally, there are placeholders in the TM match, but none in the XLIFF. The business unit is the same. The *match-quality* is set to 99.5%, *dx:glorious-match* is set to "no", and *dx:match-penalty* is set to 2%. The CAT tool displays this match to the user as a 97.5% match.
- ☀ Example 3: No match is found for a sentence in the TM, so a machine translation is applied. The "source" for the machine translation is an exact match, textually, to the source in the XLIFF. The *match-quality* is set to 100%, *dx:glorious-match* is set to "no", and *dx:match-penalty* is set to 10%. The CAT tool displays this match to the user as a 90% match.
- ☀ *Note: Processing Tools can choose to combine this information to display an overall match rate (100% text match - 3% attribute mismatch penalty = 97% overall match), or they can display it separately. This behavior is not dictated by XLIFF:doc.*

Use the *origin* and *dx:origin-shorttext* attributes to describe where the match candidate came from.

Use the XLIFF `<alt-trans><target>` element's *state* attribute to indicate the status of the proposed match (if one was assigned to it in the originating TM, etc.). See the reference section for a list of the acceptable values for this attribute.

## Providing Context for Matches

XLIFF:doc supports an option for the Creation Tool to provide a degree of textual context for embedded TM matches. This functionality can be used to supplement the information provided to the translator via *match-quality*, *dx:glorious-match*, etc. "Context," in this scenario, refers to the text that came before and after the text in the actual TU match. The Creation Tool can provide up to 150 characters (not bytes) of context for what came before the match, and up to 150 characters for what came after the match. Nothing but text is allowed in the context elements, so all tags, inlines, etc., must be resolved into plain text. Additionally, to keep XLIFF:doc file size down, and to reinforce the goal of this element being textual context for translators, Creation Tools must resolve inline placeholders into white space, words, or nothing at all. Do not use "{1}", "[1]", etc. If the segment being embedded as a match did not have any other text before or after it, the Creation Tool may choose to substitute other information. For example, "[XML structure]", "[Start of File]", "[End of File]", etc. However, this behavior is not defined by XLIFF:doc, except to the extent that it must of course contain only text.

- ☀ *Note: goal of this functionality is to allow TM engines that already provide textual context for edit sessions done in the tool's own translation editor, to be able to provide at least some text to users who are editing in a different translation editor, via XLIFF:doc. Embedding this information is optional for all Creation Tools.*

Let's look at an example. Assume the translator is translating the following section from a medical device technical manual, and is currently translating the third bullet:

- ☀ Example:

### **4 Warnings and precautions**

*Inspecting the sterile package – Inspect the sterile package with care before opening it.*

- *Contact an Acme Corp. representative if the seal or package is damaged.*

## Interoperability

- Do not store this product above 40 °C.
- Do not use the product after its expiration date.

**Sterilization** – Acme Corp. has sterilized the package contents with ethylene oxide before shipment. This lead is for single use only and is not intended to be resterilized.

In our example, let's say the translator has an exact text match from a TM, for the third bullet, "Do not use the product after its expiration date," but the match is not an in-context ("*dx:glorious-match=no*") match. And let's also say that the translator is translating into Japanese, but not for a medical device, for but a food product. The translator is provided with a match that looks good, but they want to see if the translation is appropriate for their current context. Using the context provided by the Creation Tool (shown in green), the translator's Processing Tool (editor) provides this merged view of the before-current-after data:

Crispy donuts are delicious, healthy, and part of a balanced daily meal. Do not use the product after its expiration date. Unless you are recovering from a hang-over, in which case you will probably only notice the sugar anyway.

The translator can quickly see from the context information that the embedded match is for a food product, and not for a medical translation, which perhaps means they will decide to edit the translation match.

- ☼ Note: It is optional for the Creation Tool to provide this context. If provided, however, any Processing Tool must make the context information available to translators. XLIFF:doc sets no expectations or requirements about what mechanism a Processing Tool uses to show embedded context to users.
- ☼ Note: Context, if provided, must be provided for both the source and target.

## Describing Attributes of the Match

Use the XLIFF:doc elements `<dx:attributes>` and `<dx:attribute>` to provide meta information to the user and Processing Tool about the match candidates. This could include user definable pairs (eg, "Product", "Company", "Project", etc.), as well as system-defined information such as the username of the person who last edited the translation, timestamp for the last edit, etc.

## Usage Example

```
<alt-trans alttranstype="proposal" state="signed-off" match-quality="100" origin="TM"
dx:origin-shorttext="Master TM 1" dx:modified-by="smithj1" dx:modified-at="20040805T00:12:34Z">
  <source>The potential adverse events related to the use of transvenous leads include, but are
not limited to, the following patient-related conditions:</source>
  <target>Parmi les effets indésirables potentiels associés à l'utilisation de sondes endocavi-
taires on peut citer, entres autres, les complications patient suivantes :</target>
  <dx:note dx:modified-by="Jane Doe" dx:modified-at="2010-10-12T00:00:00Z">This is a really old
translation!</dx:note>
  <dx:context>
    <dx:context-source-before>Potential adverse events</dx:context-source-before>
    <dx:context-source-after>• air embolism • cardiac dissection • cardiac
perforation</dx:context-source-after>
    <dx:context-target-before>Effets secondaires potentiels</dx:context-target-before>
    <dx:context-target-after>• embolie gazeuse • dissection cardiaque • perforation
cardiaque</dx:context-target-after>
  </dx:context>
  <dx:attributes>
    <dx:attribute name="Company">Medtronic, Inc.</dx:attribute>
    <dx:attribute name="Business Unit">CRDM</dx:attribute>
    <dx:attribute name="Domain">Therapy Delivery</dx:attribute>
    <dx:attribute name="Audience">clinician</dx:attribute>
    <dx:attribute name="Project ID">GTSID001990,GTSID002968,GTSID003131</dx:attribute>
  </dx:attributes>
</alt-trans>

<alt-trans alttranstype="proposal" match-quality="100" origin="MT" dx:match-penalty="15"
dx:origin-shorttext="MT Engine 2" dx:modified-by="mtengine2" dx:modified-at="20101013T00:00:00Z">
```


## Interoperability

`<source>` The potential adverse events related to the use of transvenous leads include, but are not limited to, the following patient-related conditions:`</source>`  
`<target>`Les événements indésirables potentiels liés à l'utilisation de sondes endocavitaires incluent, mais ne sont pas limités à, les suivants liés au patient conditions suivantes:`</target>`  
`</alt-trans>`

### Terminology Matches

XLIFF:doc supports the ability for any tool in the process to embed terminology data specific to the project. In practice, this would likely be done by the Creation Tool.

XLIFF:doc supports the concept of TU-specific terminology. In contrast to the traditional model of simply providing a termbase and expecting the translation editor application to find the appropriate matches from that termbase, XLIFF:doc pre-embeds the appropriate term matches for each TU. As a result, it does not matter how the translation editor application analyzes and matches terminology against a source string, the pre-defined terms will always be shown. In fact, the translation editor does not have to know anything about stemming terms, or how to read a TBX file. All it needs to do is read the IDs from the pre-embedded term candidates for each TU, look the IDs up against the included termbase, and show the hits to the end user. This not only makes it easier for tools developers to support terminology in translation editing tools, but it gives more control to the owner of the terminology: the “right” terms will always be shown. In theory, terminology can be embedded by any tool in the project workflow. In practice, it is most likely going to be the job of the XLIFF Creation Tool.

 *Note: The ability to embed terminology does not mean a translation tool cannot also take advantage of a general termbase. The Translation Interoperability Protocol (TIP) Package provides for TBX files to be included in the translation package, alongside the XLIFF. While compatible tools are required to process and display any embedded term matches present within the XLIFF, there is no requirement that matches actually be embedded by the Creation Tool. Additionally, a tool could display term matches from multiple sources (a term server, a TBX, etc.), not just the embedded terms.*

### Term Embedding Mechanism


XLIFF:doc uses a glossary and a look-up model to embed terminology data. A UTX-format glossary consisting only of terms referenced in the XLIFF:doc is embedded into a `<dx:utx-glossary>` element. Each TU with a relevant term candidate has one or more `<mrk>` elements.

### Embedding Term Matches


To embed terminology data in an XLIFF:doc file, the general process might be something like this:

Read terminology source (a termbase file, data from a terminology server, a csv file, etc.).

- Apply filter(s) as necessary
- Match terminology against each TU in the XLIFF, applying filter(s) as necessary/available.
- Insert all matched terms into the `<dx:utx-glossary>` element (under the `<xliff>` root element).

 *Note: Do not include the same term more than once in the glossary! You can reference the same term as many times as required, by using the `<mrk>` element as needed.*

- If the termbase is concept-based, and you want to make synonyms within a matched concept available to the end user, enter each synonym in a different `<dx:utx-term>` element in the glossary.

 *Note: It is the responsibility of the Creation Tool to pick the best terminology match. If there are synonyms, it is the responsibility of the Creation Tool to do the "right thing". That may mean using the meta information available to it to pick the most appropriate synonym. It may mean including all synonyms, and giving the Processing Tool the responsibility for displaying synonyms. This behavior is not dictated by XLIFF:doc. XLIFF:doc does require all Processing Tools to make all referenced (via `<mrk>`) term matches embedded in the file available to users.*



## Interoperability

- For each source match found, encapsulate the matched text in the `<source>` in a `<mrk>` element. Reference the ID of the term (as listed in the glossary) via the `dx:term-id` attribute of the `<mrk>` element.
- Optionally: if target text is available when creating the XLIFF:doc, and a term match is found in the target, encapsulate the matched text in the `<target>` in a `<mrk>` element. Reference the ID of the term (as listed in the glossary) via the `dx:term-id` attribute of the `<mrk>` element.

☼ *Note: it is not the responsibility of the Creation Tool to determine whether identified terms are being used correctly in the XLIFF:doc. Accordingly, there are no provisions for linking `<mrk>` elements in the source to `<mrk>` elements in the target.*

## Mapping TBX to UTX

XLIFF:doc uses the [Universal Terminology Exchange \(UTX\)](#) format for embedded terms. UTX is a simple tab-delineated format that provides easy access to basic terminology information. Before the UTX can be embedded in XLIFF:doc, it needs to be mapped to XLIFF:doc xml elements. The table below will help you map between TBX, UTX, and XLIFF:doc UTX elements.

TBX	UTX before XLIFF:doc conversion	<dx:utx-glossary>
termEntry	[a line in the UTX glossary]	<dx:utx-term>
termEntry@id	"concept ID" column	dx:utx-term@dx:concept-id
termEntry/langSet	Use to map the termEntry/langSet/tig/term to either the "src" or "tgt" column	dx:utx-term@dx:src-lang and dx:utx-term@dx:tgt-lang
termEntry/langSet/tig/term	"src" or "tgt" column	dx:utx-src or dx:utx-tgt
termEntry/langSet/tig/termNote	"comment" column	dx:utx-comment
termEntry/langSet/tig/termNote@type	n/a	n/a
termEntry/langSet/tig/termDescrip	n/a (or map to "comment" column in place of termNote)	n/a or map to dx:utx-comment
termEntry/langSet/tig/termDescrip@type	n/a	n/a

## UTX as Supported by XLIFF:doc

In order to provide interoperability, XLIFF:doc defines the following requirements for embedded UTX glossaries:

- The UTX must be mapped to the `<dx:utx-glossary>` element and its children, as defined in XLIFF:doc. It is not valid to include a UTX as a block of tab-separated text.
- XLIFF:doc requires an 8 column UTX: src, tgt, src:pos, term status, comment, concept ID, entry ID.
- The "entry ID" column value must be unique within the UTX file. This is the value that will be used in the `<mrk>` `dx:term-id` attribute, to generate a reference from the TU to the glossary.
- The value of the "term status" column must match the status of the translation entry, not the source entry.
- XLIFF:doc does not specify what a tool must do with the src:pos, comment, or concept ID values, other than make them available to the user.

☼ *Note: One possible use for the concept ID value is to refer back to a TBX glossary included as reference in a TIP package.*

## Usage Example

```
<trans-unit id="10">
  <source>The <mrk mtype="term" dx:term-id="0">Commodore 64</mrk> is the best-selling single per-
sonal computer model of all time. </source>
  <target state="new"></target>
```

## Interoperability

```
</trans-unit>
<trans-unit id="11">
  <source>The C-64 or <mrk mtype="term" dx:term-id="1">C=64</mrk> as it was known to fans had
impressive graphics at an affordable price. </source>
  <target state="new"></target>
</trans-unit>
<trans-unit id="12">
  <source>Its successor, the Commodore <mrk mtype="term" dx:term-id="2">Amiga</mrk>, amazed the
computing world when it was launched, but was never able to capture the market share enjoyed by
the <mrk mtype="term" dx:term-id="0">Commodore 64</mrk>. </source>
  <target state="new"></target>
</trans-unit>
```

... (dx:utx-glossary is a child of the <xliiff> element.) ...

```
<dx:utx-glossary dx:version="1.11" dx:src-lang="en-US" dx:tgt-lang="ja-JP"
dx:date-created="2011-08-17T19:00:00Z" dx:creator="Acme Corporation">
  <dx:utx-term dx:entry-id="0" dx:concept-id="150">
    <dx:utx-src dx:pos="noun">Commodore 64</dx:utx-src>
    <dx:utx-tgt dx:status="approved">コモドール64</dx:utx-tgt>
    <dx:utx-comment>The Commodore 64 is an 8-bit home computer introduced by Commodore Interna-
tional in January 1982.</dx:utx-comment>
  </dx:utx-term>
  <dx:utx-term dx:entry-id="1" dx:concept-id="150">
    <dx:utx-src dx:pos="noun">C=64</dx:utx-src>
    <dx:utx-tgt dx:status="forbidden">C=64</dx:utx-tgt>
    <dx:utx-comment>The C=64 had a terrible keyboard (unless, of course you compare it to an IBM
PCjr keyboard).</dx:utx-comment>
  </dx:utx-term>
  <dx:utx-term dx:entry-id="2" dx:concept-id="172">
    <dx:utx-src dx:pos="noun">Amiga</dx:utx-src>
    <dx:utx-tgt dx:status="approved">アミーガ</dx:utx-tgt>
    <dx:utx-comment>The Amiga is a family of personal computers sold by Commodore in the 1980s
and 1990s. The first model was launched in 1985 as a high-end home computer and became popular
for its impressive graphical, audio and multi-tasking abilities.</dx:utx-comment>
  </dx:utx-term>
  <dx:utx-term dx:entry-id="3" dx:concept-id="172">
    <dx:utx-src dx:pos="noun">Amiga</dx:utx-src>
    <dx:utx-tgt dx:status="non-standard">Amiga</dx:utx-tgt>
    <dx:utx-comment>The Amiga is a family of personal computers sold by Commodore in the 1980s
and 1990s. The first model was launched in 1985 as a high-end home computer and became popular
for its impressive graphical, audio and multi-tasking abilities.</dx:utx-comment>
  </dx:utx-term>
</dx:utx-glossary>
```

## Reading and Displaying Embedded Term Matches

XLIFF:doc does not require any particular behavior on the part of the tools reading the terms, other than to make them available to the user. Because the XLIFF:doc terminology markup scheme uses <mrk>, and because the <mrk> tag points to a specific term entry (not a term concept), it is the responsibility of the Processing Tool to identify any synonyms and show them to users, if that functionality is desired. The embedded glossary's *dx:concept-id* attribute can be used to identify related terms (synonyms, short forms, etc.).

## QA Information

XLIFF:doc allows for any tool in the chain to add and remove QA messages to the XLIFF file. QA messages are applied to the `<trans-unit>` element as a whole, because a QA issue often involves both the source and the target. Each TU can have at most one `<dx:qa-hits>` element, but each `<dx:qa-hits>` element can have as many `<dx:qa-hit>` elements as necessary.

☼ *Note: XLIFF:doc does not specify the logic a Creation Tool uses to identify errors and generate QA messages, or what a Processing Tool does with the QA messages.*

If the QA hit can be traced to a particular spot in the source or target (or both), the position of the first character of the affected text, and the length of the affected text, can be placed in these attributes to provide the Processing Tool an opportunity to highlight, link, or otherwise indicate to the user the text related to the QA hit. The `dx:qa-tgt-startpos` (or `dx:qa-src-startpos`) is the position from the start of the segment, with 1 as the first position. The `dx:qa-tgt-length` is the length in characters of the affected text. If the QA hit is related to text in both the source and text (`dx:qa-origin="both"`), then both pairs of attributes must be filled out. Both the `dx:qa-tgt-startpos` and `dx:qa-tgt-length` attributes must be filled out. The Processing Tool must not attempt to visually highlight the QA hit text, if either or both of these attributes are set to "0".

Beyond this, XLIFF:doc does not specify what a Processing Tool should do with this information.

☼ *Note: `dx:qa-tgt-startpos` and `dx:qa-tgt-length` are both 1-based integers, and both refer to Unicode characters, not necessarily bytes.*

☼ *Note: Processing Tools are not to perform any Unicode normalization on text. Normalization can result in invalidating the position markers from tool to tool. See [Unicode Standard Annex #15: Unicode Normalization Forms](#) for more information.*

☼ *Note: when specifying start position and length, you must account for tags when setting the start position and length: add 1 character for each tag. See the second example below.*

☼ *Note: if the translation is modified by a tool or by a user at any point after the QA information is embedded, the QA information may no longer be accurate. The position of the text referred to by the QA hit may have moved, the text may have been deleted, etc. QA hits are less permanent in nature than TM and terminology matches, and Processing Tools should take that into consideration when processing and displaying them.*

If the XLIFF:doc flags set by the Creation Tool allow it, tools must give users the ability to remove QA hits the user no longer wishes to see. See the Reference section for more information.

### Usage Example 1

Source: The pump internal volume (internal tubing volume) for all models of the pump is 300 µL.

Target: Het interne pompslangvolume bedraagt voor alle pompmodellen 300 ml.

Resulting Markup:

```
<dx:qa-hits>
  <dx:qa-hit dx:qa-origin="source" dx:qa-code="005" dx:qa-ignorable="no" dx:qa-shorttext="Number
  found in source, not found in target: &quot;300 µl&quot;" dx:qa-category="number check"
  dx:qa-level="error" dx:qa-src-startpos="81" dx:qa-src-length="6" />
  <dx:qa-hit dx:qa-origin="target" dx:qa-code="005" dx:qa-ignorable="no" dx:qa-shorttext="Number
  found in target, not found in source: &quot;300 ml&quot;" dx:qa-category="number check"
  dx:qa-level="error" dx:qa-tgt-startpos="61" dx:qa-tgt-length="6" />
</dx:qa-hits>
```

☼ *Note: the list of acceptable values for the `dx:qa-category` attribute can be found in the Reference Section*

### Usage Example 2

Source: `<g id="1" ctype="x-other" dx:orig-markup-open="&lt;acronym&gt;" dx:orig-markup-close="&lt;/acronym&gt;">IEC</g> 601-1-2.`

Target: `<g id="1" ctype="x-other" dx:orig-markup-open="&lt;acronym&gt;" dx:orig-markup-close="&lt;/acronym&gt;">IEC</g> 601-1-1`

Resulting Markup:

```
<dx:qa-hits>
  <dx:qa-hit dx:qa-origin="source" dx:qa-code="025" dx:qa-ignorable="no" dx:qa-shorttext="IEC
    Standard found in source, not found in target: &quot;IEC 601-1-2&quot;;"
    dx:qa-category="pattern" dx:qa-level="error" dx:qa-src-startpos="2" dx:qa-src-length="13" />
  <dx:qa-hit dx:qa-origin="target" dx:qa-code="025" dx:qa-ignorable="no" dx:qa-shorttext="IEC
    Standard found in target, not found in source: &quot;IEC 601-1-1&quot;;"
    dx:qa-category="pattern" dx:qa-level="error" dx:qa-tgt-startpos="2" dx:qa-src-length="12" />
  <dx:qa-hit dx:qa-origin="source" dx:qa-code="031" dx:qa-ignorable="no" dx:qa-shorttext="End of
    sentence punctuation in source different from target: &quot;.&quot;;"
    dx:qa-category="punctuation" dx:qa-level="warning" dx:qa-src-startpos="12"
    dx:qa-src-length="1" />
  <dx:qa-hit dx:qa-origin="source" dx:qa-code="029" dx:qa-ignorable="yes" dx:qa-shorttext="Two
    or more continuous spaces in source" dx:qa-category="punctuation" dx:qa-level="warning"
    dx:qa-src-startpos="4" dx:qa-src-length="2" />
</dx:qa-hits>
```

## Revision History

XLIFF:doc provides a mechanism for Creation Tools to embed version history for each TU in an XLIFF:doc. This version history can then be displayed by any Processing Tool. In XLIFF:doc, version history is stored on the translation unit basis, and is read-only. A Processing Tool would not record new version history as a user edits a file. Only a Creation Tool can create version history. This reflects the design goal of XLIFF:doc, to be a temporary vessel for exchanging data between two systems, and not to be a full-blown, permanent translation bitext file format that knows about every last thing that every happened for a project or file.

☀ *Note: while adding version history to an XLIFF:doc is optional for the Creation Tool, if present, version history must be shown to users by the Processing Tool.*

## Unit Histories

Past versions of individual translation units are placed using the `<alt-trans>` element, with the `alttranstype` attribute set to "previous-version". The following pieces of TU history can be embedded:

- Who created this particular version of a TU (`dx:modified-by`)
- When this version was created (`dx:modified-at`)
- What the translated text was before it was updated
- The origins, quality, and match penalty (if any) associated with the TU before it was updated
- The status of the target translation before it was updated

If a Creation Tool has access to version history for a TU in a project, and if the Creation Tool is set up to add version history to XLIFF:docs, it would add one `<alt-trans>` element for each version of the translation unit it knows about (excluding the current version, which is already represented in the main `<trans-unit>`).

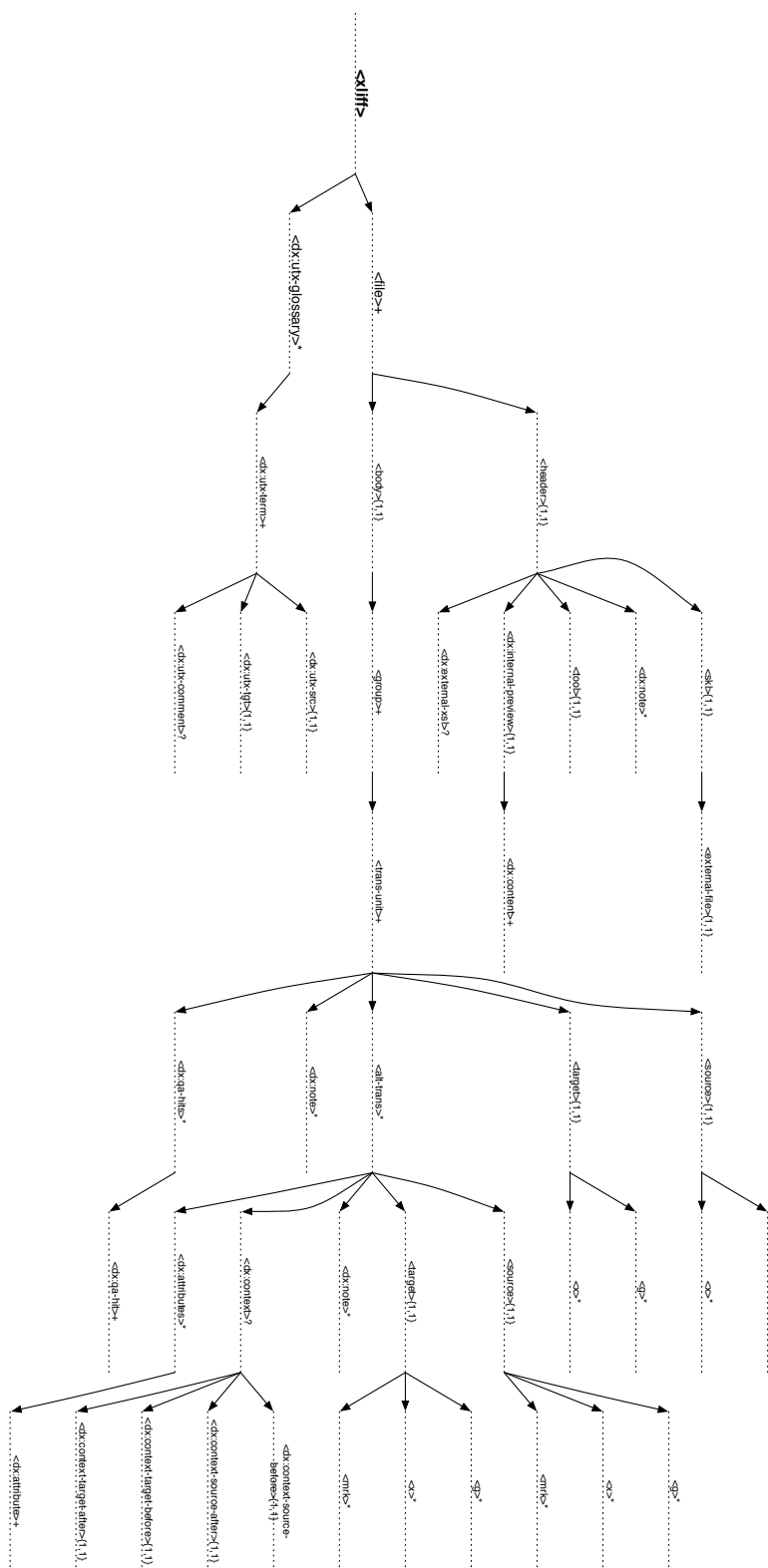
☀ *Note: Do not create new a `<alt-trans>` element if the translation is the same as the immediately proceeding `<alt-trans>`. XLIFF:doc does allow for `<alt-trans>` elements to have identical `<target>` contents, but only if they are non-sequential. For example, the first translation (version 1) was edited by a reviewer (version 2), and then a second reviewer changed it back (version 3) to the original version. In this case, you could have three `<alt-trans>`, with the `<target>` of `<alt-trans>` 1 and 3 being identical.*

☀ *Note: Revision history is only captured for the target, since the source is immutable in XLIFF:doc.*

## Usage Example

```
<trans-unit>
...
  <alt-trans alttranstype="previous-version" dx:modified-by="l10n-tool"
dx:modified-at="2011-03-04T02:30:00Z" dx:match-quality="100" dx:match-penalty="99" dx:origin="MT"
dx:origin-shorttext="PseudoTrans">
    <target state="new">XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX
XXXXX</target>
  </alt-trans>
  <alt-trans dx:modified-by="ElizaTranslator" dx:modified-at="2011-03-04T03:30:00Z"
dx:match-quality="100" dx:match-penalty="15" dx:origin="MT" dx:origin-shorttext="Google MT">
    <target state="translated">Quant à la reléguer dans le domaine de la fiction, cette charge
ont dû être abandonnées.</target>
  </alt-trans>
</trans-unit>
```

## Appendix A: Diagram of XLIFF:doc Elements



## Appendix B: Reference Guide to XLIFF:doc

- ☼ *Note: all elements mentioned here must be processed and acted upon by all Creation Tools and Processing Tools. All required attributes must be processed. All optional attributes must be at least passed through without loss, if the tool is not going to act on the attribute. Any element or attribute not specified here must not be used.*
- ☼ *Note: By necessity, this reference includes information on the XLIFF namespace. This is not meant to be a reference to the entire XLIFF namespace, only the portion used by XLIFF:doc. Please refer to the [XLIFF 1.2 specifications](#) for more information.*
- ☼ *Note: this reference section uses regex-inspired conventions to denote how many child elements a particular XML element should have. The following examples may help:*
  - `{1,1}`    a single child element is required
  - `{1,3}`    at least one child element is required, no more than 3
  - `+`        at least one child element is required (no upper limit)
  - `?`        optionally, one child element
  - `*`        optionally, one or more child elements

### <alt-trans>

In XLIFF:doc, alt-trans is used for 2 purposes. Requirements are different depending on usage:

1. Embedded translation matches from TM, MT, etc. Use *alttranstype* = "proposal".

☼ Example: `<alt-trans alttranstype="proposal" match-quality="99.75" dx:match-penalty="0.25" origin="TM" dx:origin-shorttext="Master TM 1" dx:modified-by="smithj1" dx:modified-at="20101012T00:00:00Z">...</alt-trans>`

2. Translation revision history. Use *alttranstype* = "previous-version". Do not include a <source> element.

☼ Example: `<alt-trans alttranstype="previous-version" dx:modified-by="JaneDoe" dx:modified-at="20101013T21:06:00Z" dx:match-quality="100" dx:match-penalty="0" dx:origin="TM" dx:origin-shorttext="MyBigTM">...</alt-trans>`

**Parent Elements:** <trans-unit>

**Child Elements:** <source>{1,1}, <target>{1,1}, <dx:note>\*, <dx:attributes>\*, <dx:context>?

**Required Attributes:** *alttranstype*, *match-quality*, *dx:glorious-match*, *dx:match-penalty*, *origin*, *dx:origin-shorttext*, *dx:modified-by*, *dx:modified-at*

*alttranstype*

Acceptable values: "proposal", "previous-version"

*match-quality*

Use values 0-100. The use of decimal numbers is acceptable, but only a single digit of precision is allowed. For example, "99.5" is acceptable, but "99.52" is not. Do not use a percent sign. Higher numbers refer to better matches than lower numbers.

*dx:glorious-match*

Indicates whether the match is an exact, in-context match (as determined by the tool that generated the match).

Acceptable values: "yes", "no". Default value: "no".

*origin*

## Interoperability

Acceptable values: "TM", "MT", "XLIFF", "manual", "sourcetext", "pseudo-trans". See the *dx:origin* attribute of the `<trans-unit>` element for more information on how to assign and interpret these values.

### *dx:origin-shorttext*

Human-readable string describing the match origin. This could be the name of the TM it came out of, the name of an MT engine, etc. Examples: "Master TM 1", "Google MT", "myDocument.docx.xlf".

### *dx:match-penalty*

Use values 0-100. The use of decimal numbers is acceptable, but only a single digit of precision is allowed. For example, "0.5" is acceptable, but "0.48" is not. Do not use a percent sign. Higher numbers refer to larger penalties than lower numbers.

### *dx:modified-by*

String value identifying the user (human or automatic) responsible for the most recent modification.

### *dx:modified-at*

Time and date the translation was last modified. This can represent the date/time it was created, editing, or processed in some other way.

☀ *Note: Date/time must be specified in ISO 8601 format. Time must be specified in UTC (Zulu) time zone. [See the XLIFF date element for more information.](#)*

**Optional Attributes:** none

## `<body>`

**Parent Elements:** `<file>`

**Child Elements:** `<group>+`

**Required Attributes:** none

**Optional Attributes:** none

## `<dx:attribute>`

Provides key-value style meta information about the parent element. Use the *dx:name* element for the key, and place the value into the text child.

**Parent Elements:** `<dx:attributes>`

**Child Elements:** none

**Required Attributes:** *dx:name*

*name*

The "key" part of the attribute key/value pair. May be used by a Processing Tool to display the attribute to the user.

`<dx:attributes>`

```
<dx:attribute dx:name="Company">Acme, Inc.</dx:attribute>
<dx:attribute dx:name="Business Unit">Advanced Rockets and Anvils</dx:attribute>
</dx:attributes>
```

**Optional Attributes:** none

## `<dx:attributes>`

Container for holding individual attributes (`<dx:attribute>`) associated with the parent object.



## Interoperability

**Parent Elements:** [<alt-trans>](#)

**Child Elements:** [<dx:attribute>+](#)

**Required Attributes:** none

**Optional Attributes:** none

### [<dx:content>](#)

A child-less element that marks the place in a preview element or file where a translation unit's source or target should be placed to create a preview.

**Parent Elements:** [<dx:internal-preview>](#)

**Child Elements:** none

**Required Attributes:** *id*

*id*

Reference to the *id* attribute of the [<trans-unit>](#) whose [<source>](#) or [<target>](#) content will be substituted in place of the [<dx:content>](#) element in the preview.

**Optional Attributes:** none

### [<dx:context>](#)

Container for holding individual pieces of textual context associate with the parent match candidate.

**Parent Elements:** [<alt-trans>](#)

**Child Elements:** [<dx:context-source-before>\\*](#), [<dx:context-source-after>\\*](#), [<dx:context-target-before>\\*](#), [<dx:context-target-after>\\*](#)

**Required Attributes:** none

**Optional Attributes:** none

### [<dx:context-source-after>](#)

The contents of this element are a text-only representation of the context that came after the content captured in the [<source>](#) of the [<alt-trans>](#) ancestor. A maximum of 150 characters (not bytes) worth of information can be stored in this element.

**Parent Elements:** [<dx:context>](#)

**Child Elements:** none

**Required Attributes:** none

**Optional Attributes:** none

### [<dx:context-source-before>](#)

The contents of this element are a text-only representation of the context that came before the content captured in the [<source>](#) of the [<alt-trans>](#) ancestor. A maximum of 150 characters (not bytes) worth of information can be stored in this element.

**Parent Elements:** [<dx:context>](#)

**Child Elements:** none

## Interoperability

**Required Attributes:** none

**Optional Attributes:** none

### <dx:context-target-after>

The contents of this element are a text-only representation of the context that came after the content captured in the <target> of the <alt-trans> ancestor. A maximum of 150 characters (not bytes) worth of information can be stored in this element.

**Parent Elements:** <dx:context>

**Child Elements:** none

**Required Attributes:** none

**Optional Attributes:** none

### <dx:context-target-before>

The contents of this element are a text-only representation of the context that came before the content captured in the <target> of the <alt-trans> ancestor. A maximum of 150 characters (not bytes) worth of information can be stored in this element.

**Parent Elements:** <dx:context>

**Child Elements:** none

**Required Attributes:** none

**Optional Attributes:** none

### <dx:external-xsl>

The content of this element must be a URL to the XSL code that will be used to generate a source or target preview. The URL must be a relative path to a local file, as XLIFF:doc requires that the target of the URL be available to all tools that process the XLIFF. The XSL code must not be escaped.

**Parent Elements:** <header>

**Child Elements:** none

**Required Attributes:** none

**Optional Attributes:** none

### <dx:internal-preview>

The content of this element is the HTML code that will be used to generate a source or target preview. The HTML code must be escaped. The preview code must include one <dx:content> element for each <trans-unit> in each <file> element.

**Parent Elements:** <header>

**Child Elements:** <dx:content>+

**Required Attributes:** none

**Optional Attributes:** none

## Interoperability

### <dx:qa-hit>

**Parent Elements:** <dx:qa-hits>

**Child Elements:** none

**Required Attributes:** *dx:qa-origin*, *dx:qa-ignorable*, *x:qa-shorttext*, *dx:qa-category*, *dx:qa-level*

#### *dx:qa-origin*

Indicates whether the QA hit originates from the source, the target, or from both. If the QA system identifies not just errors and warnings, but also non-errors (the correct thing was done by the linguist), then it can use a value of "both" here, and collapse the source-to-target hit, and the target-to-source hit into a single hit.

Accepted values: "source", "target", "both"

#### *dx:qa-ignorable*

Provides information to the tool reading the file as to whether this QA message can safely be ignored by the user. This determination is made by the tool that creates the QA message. The information must be made available to the tool user, but further behavior is not specified or required.

Accepted values: "yes", "no"

#### *x:qa-shorttext*

Human-readable description of the QA message. Must include specific information whenever possible, rather than a general description of the error.

#### *dx:qa-category*

Categorizes the type of QA message involved. The following table may be of assistance in matching your tool's QA categories to those supported in XLIFF:doc:

Category	Description
date	A date present in one language was missing or different in the other language
inconsistency	Two or more TUs have the same source but different translations; two or more TUs have the same translation but different sources
internationalization	Corrupt characters; spacing before units of measurement; formatting of numbers does not match the standards for the target language; punctuation marks were not properly translated into the target language; translation is incorrect for other cultural (non-linguistic) reasons
number	A number found in the source is different or missing in the translation; a number found in the translation is missing or different in the source; a number is found in both source and translation, but the units of measurement do not match.
omission	Empty translations; incomplete translations
other	Significant difference in length of source text compared to translation; every other type of QA hit not covered by one of the specific categories
pattern	A regular-expression pattern or other pattern (as defined in the QA checker) found a match (or failed to find a match, depending on usage)
punctuation	Differences in spacing, punctuation, brackets, tab characters, or capitalization
tags	One or more tags (placeholders/inline formatting) is missing or different in either the source or target; one or more tags is duplicated in the translation

## Interoperability

Category	Description
terminology	Terminology usage in either the source or translation is not consistent with the termbase; terms were used that were flagged in the termbase as not-for-use
time	A time of day present in one language was missing or different in the other language

Acceptable values: "number", "pattern", "punctuation", "tags", "date", "time", "internationalization", "omission", "inconsistency", "terminology", "other"

### *dx:qa-level*

Defines the degree of seriousness of the QA message.

Accepted values: "error" (most serious), "warning", "non-error" (proactive match/confirmation)

**Optional Attributes:** *dx:qa-code*, *dx:qa-tgt-startpos*, *dx:qa-tgt-length*, *dx:qa-src-startpos*, *dx:qa-src-length*

### *dx:qa-code*

Tool-specific ID code for the error type. XLIFF:doc does not specify any limitations for the value of this optional attribute.

☼ *Note: this is not an ID number for the specific QA hit instance. There is no ID attribute in XLIFF:doc for a QA hit instance.*

### *dx:qa-tgt-startpos*

The position, in unicode characters, from the start of the target segment, with 1 as the first position. Set to "0" if the position is unknown.

### *dx:qa-tgt-length*

The length, in unicode characters, of the text in the target related to the QA hit. Set to "0" if the length is unknown.

### *dx:qa-src-startpos*

The position, in unicode characters, from the start of the source segment, with 1 as the first position. Set to "0" if the position is unknown.

### *dx:qa-src-length*

The length, in unicode characters, of the text in the source related to the QA hit. Set to "0" if the length is unknown.

## <dx:note>

In XLIFF:doc, note objects can be applied to the file itself (via the <header> element), and translation pairs (via the <trans-unit> and <alt-trans> elements).

☼ **Example:**

```
<dx:note dx:modified-by="JoeTranslator" dx:annotates="source" dx:modified-at="2011-01-14T09:30:24Z">"driver" = typo in the source? The reference material shows a cat in a scuba suit, not in a race car.</dx:note>
<dx:note dx:modified-by="-JaneContentAuthor" dx:annotates="source" dx:modified-at="2011-01-14T11:42:43Z">Good catch, thanks. Will fix to read "diver" in next revision.</dx:note>
```

**Parent Elements:** <trans-unit>, <alt-trans>, <header>

**Child Elements:** none

**Required Attributes:** *dx:modified-at*, *dx:modified-by*

## Interoperability

### *dx:modified-at*

Time and date the `<dx:note>` was last modified. This can represent the date/time the note was created, edited, or processed in some other way.

☀ *Note: Date/time must be specified in ISO 8601 format. Time must be specified in UTC (Zulu) time zone. [See the XLIFF date element for more information.](#)*

### *dx:modified-by*

Use this attribute to indicate who created or last edited the note. This would typically be a user ID, user name, role or description.

Acceptable values: undefined. Must be a string value.

**Optional Attributes:** *dx:annotates*

### *dx:annotates*

Notes can optionally be defined as referring to specific text in the source or target. If specific text is to be associated with the note, the *dx:annotates* attribute must be set, and set to either source or target.

Acceptable values: source, target, general. Default value: general.

## `<dx:qa-hits>`

Container for individual QA hit elements.

**Parent Elements:** `<trans-unit>`

**Child Elements:** `<dx:qa-hit>+`

**Required Attributes:** none

**Optional Attributes:** none

## `<dx:utx-comment>`

This element maps optional comment information from a UTX glossary. This information must be presented to the user. What you put into the comment, or whether you include one at all, is not defined by XLIFF:doc.

**Parent Elements:** `<dx:utx-term>`

**Child Elements:** none

**Required Attributes:** none

**Optional Attributes:** none

## `<dx:utx-glossary>`

This element contains data from a UTX glossary, mapped to XML. The contents of this element will be referenced by individual terminology hits within each TU.

☀ *Note: A UTX glossary in an XLIFF:doc file may only contain languages that match the languages of the XLIFF:doc itself. For example, an English to French XLIFF:doc cannot contain a UTX glossary that references German and English. However, it is permissible for a tool to reference languages that are only a match on the language, and not on the geography/culture of the XLIFF:doc. For example, an XLIFF:doc going into French for France (fr-FR) could have an embedded UTX glossary that referenced French for Canada (fr-CA) terms.*

## Interoperability



Example:

```
<dx:utx-glossary dx:version="1.11" dx:src-lang="en-US" dx:tgt-lang="ja-JP" dx:date-created="2011-08-17T19:00:00Z" dx:creator="Acme Corporation">...</dx:utx-glossary>
```

**Parent Elements:** <xliif>

**Child Elements:** <dx:utx-term>+

**Required Attributes:** *dx:version*, *dx:src-lang*, *dx:tgt-lang*, *dx:date-created*, *dx:creator*

*dx:version*

Maps to the version item in the header of the UTX file (position 1)

*dx:src-lang*

Maps to the source language item in the header of the UTX file (position 2 - before slash). ISO 639 and 3166 formats.



*Note:* See [RFC 5646](#) for more information.

*dx:tgt-lang*

Maps to the target language item in the header of the UTX file (position 2 - after slash). ISO 639 and 3166 formats.



*Note:* See [RFC 5646](#) for more information.

*dx:date-created*

Maps to the date created timestamp in the header of the UTX file (position 3). ISO 8601 format.

*dx:creator*

Maps to the creator item in the header of the UTX file (position 4)

**Optional Attributes:** none

### <dx:utx-src>

This element maps information related to the source language, from a UTX glossary (column 1). The text contents of the element must consist of the term as represented in the source language.

**Parent Elements:** <dx:utx-term>

**Child Elements:** none

**Required Attributes:** *dx:pos*

*dx:pos*

This is the source part-of-speech from the UTX glossary (column 3). See the UTX standard for more information.

Acceptable values: noun, properNoun, verb, adjective, adverb, sentence

**Optional Attributes:** none

### <dx:utx-term>

This element has one function: to provide a look-up reference to the embedded UTX glossary. It must contain 1 <dx:utx-src> element and 1 <dx:utx-tgt> element. It may optionally have 1 <dx:utx-comment> element.

**Parent Elements:** <dx:utx-glossary>

**Child Elements:** <dx:utx-src>{1,1}, <dx:utx-tgt>{1,1}, <dx:utx-comment>?

**Required Attributes:** *dx:entry-id*

*dx:entry-id*

## Interoperability

This is a unique id that will be referenced by every TU that has a hit for this term.

Acceptable values: any value that is unique within the XLIFF:doc file.

**Optional Attributes:** *dx:concept-id*

*dx:concept-id*

This is an optional reference to an external glossary. XLIFF:doc does not require a tool to act on this id if provided, by, for example, providing lookup against external terminology sources, such as TBX files included in the TIP, a term server, etc.

### <dx:utx-tgt>

This element maps information related to the target language, from a UTX glossary (column 2). The text contents of the element must consist of the term as represented in the target language.

**Parent Elements:** <dx:utx-term>

**Child Elements:** none

**Required Attributes:** *dx:status*

*dx:status*

This is the status of the target language term, from the UTX glossary (optional column in UTX, not optional in XLIFF:doc). See the UTX standard for more information.

Acceptable values: provisional, approved, non-standard, forbidden

**Optional Attributes:** none

### <external-file>

The content of this element must be a URL to the skeleton file containing the original file's structural code. This can be an absolute path to a file on a server, a relative path to a local file, etc. This element must be specified in XLIFF:doc.

XLIFF:doc does not, however, require that the target of the URL be available to all tools that process the XLIFF. This gives a Creation Tool freedom to include the skl file in the file package with the XLIFF where any user of the XLIFF can access it, or to leave it on the server where it can only be accessed by authorized users.

**Parent Elements:** <skl>

**Child Elements:** none

**Required Attributes:** none

**Optional Attributes:** none

### <file>

**Parent Elements:** <xliff>

**Child Elements:** <header>{1,1}, <body>{1,1}

**Required Attributes:** *original*, *source-language*, *datatype*, *target-language*, *xml:space*, *dx:allow-tu-join*, *dx:allow-tu-split*, *dx:allow-tm-match-addition*, *dx:allow-tm-match-deletion*, *dx:allow-tb-match-addition*, *dx:allow-tb-match-deletion*, *dx:allow-qa-addition*, *dx:allow-qa-deletion*

*xml:space*

Controls whether whitespace is preserved for all TUs under this <file>. For more information [see the section on xml:space in the XML specification](#).

## Interoperability

Acceptable values: "default", "preserve"

☀ *Note: `xml:space` is only allowed in the `<file>` element in XLIFF:doc.*

### *original*

Reference to the original source file. Details of specification are not defined by XLIFF:doc, and can be decided on by the Creation Tool.

Acceptable values: not defined, but must be a string.

### *datatype*

This attribute specifies the kind of text contained in the `<file>` element. XLIFF:doc does not define a preset list of values for this attribute: the values are extensible. Use the following pattern when specifying this value: "x-[mimetype]". The value should reflect the mimetype defined for this type of file. For example, "x-application/vnd.oasis.opendocument.text".

### *source-language, target-language*

Acceptable values: as defined by ISO 639-1 (2-letter codes) or ISO 639-2T (3-letter codes). The use of a country code extension (based on ISO 3166) to the language code is acceptable. If extensions are used, they must use the dash "-" character, not the underscore. eg, "source-language="en-US". Lower case language codes and upper case country codes are recommended, but all tools must treat the values as case insensitive.

☀ *Note: See [RFC 5646](#) for more information.*

### *dx:allow-tu-join*

Determines whether or not TU joining is allowed in principle, for the file in question. Actual join ability may be further limited by document structure. Processing Tools are not permitted to modify this value. Default value: "yes".

### *dx:allow-tu-split*

Determines whether or not TU splitting is allowed for the file in question. Processing Tools are not permitted to modify this value. Default value: "yes".

### *dx:allow-tm-match-addition*

Determines whether or not a Processing Tool is allowed to embed additional TM matches to the file. If this value is "no", then only the Creation Tool can embed TM matches in a file. If the value is "yes", then every compatible tool is allowed to embed its own matches. Processing Tools are not permitted to modify this value. Default value: "no".

### *dx:allow-tm-match-deletion*

Determines whether or not a Processing Tool is allowed to delete embedded TM matches from the file. If this value is "yes", then tools can allow users to delete TM matches from the XLIFF. Processing Tools are not permitted to modify this value. Default value: "no".

### *dx:allow-tb-match-addition*

Determines whether or not a Processing Tool is allowed to embed additional terminology matches to the file. If this value is "no", then only the Creation Tool can embed terminology matches in a file. If the value is "yes", then every compatible tool is allowed to embed its own matches. Processing Tools are not permitted to modify this value. Default value: "no".

### *dx:allow-tb-match-deletion*

Determines whether or not a Processing Tool is allowed to delete embedded terminology matches from the file. If this value is "yes", then tools can allow users to delete terminology matches from the XLIFF. Processing Tools are not permitted to modify this value. Default value: "no".

### *dx:allow-qa-addition*



## Interoperability

Determines whether or not a Processing Tool is allowed to embed additional QA hits to the file. If this value is "no", then only the Creation Tool can embed QA hits in a file. If the value is "yes", then every compatible tool is allowed to embed its own QA hits. Processing Tools are not permitted to modify this value. Default value: "no".

*dx:allow-qa-deletion*

Determines whether or not a Processing Tool is allowed to delete embedded QA hits from the file. If this value is "yes", then tools can allow users to delete QA hits from the XLIFF. Processing Tools are not permitted to modify this value. Default value: "no".

**Optional Attributes:** none

**<g>**

**Parent Elements:** <source>, <target>

**Child Elements:** <x>\*, <mrk>\*

**Required Attributes:** *id*, *dx:orig-markup-open*, *dx:orig-markup-close*, *dx:equiv-markup-open*, *dx:equiv-markup-close*, *ctype*

*id*

The *id* is the primary mechanism for matching placeholders from source to target. The *id* must be unique within the <trans-unit>, but ids can be recycled from trans-unit to trans-unit. An *id* used for a <g> element cannot be re-used for an <x> element within the same <trans-unit>.

Acceptable values: non-negative integers

*dx:orig-markup-open*

*dx:orig-markup-close*

If the original markup can be escaped safely and included, put it into this attribute. If the original markup cannot be safely encoded in XML, do not include this attribute.

*dx:equiv-markup-open*

*dx:equiv-markup-close*

HTML code that approximates the formatting of the original document format. For example, if the original markup was a font change that made the text italicized, the *dx:equiv-markup-open* might be "&lt;i&gt;" and the *dx:equiv-markup-close* might be "&lt;/i&gt;"

☼ *Note: if the original format is HTML, and the HTML formatting can be represented safely, it is not necessary to include dx:equiv-markup attributes. If not present, the dx:orig-markup will be used.*

*ctype*

XLIFF:doc allows the following values for *ctype* in the <g> element:

Value	Description	Original HTML Markup Allowed?
bold	Indicates a run of bolded text.	Yes
italic	Indicates a run of text in italics.	Yes
underlined	Indicates a run of underlined text.	Yes

## Interoperability

link	Indicates a run of hyper-text.	No
x-superscript	Indicates a run of superscripted text.	Yes
x-subscript	Indicates a run of subscripted text.	Yes
x-fixed	Indicates a run of fixed-width text (for displaying software code, etc.)	Yes
x-font	Indicates a change in font.	Yes
x-fontsize	Indicates a change in font-size.	Yes
x-header	Indicates header type formatting. In HTML, for example, H1, H2, H3, etc.	Yes
x-title	Indicates that the enclosed text is identified as the title of some element. For example, the title attribute of an HTML <a> element.	No
x-other-format	All character level formatting not covered by specific ctype values.	No
x-other	All other types.	No

### Optional Attributes: *equiv-text*

#### *equiv-text*

Indicates the equivalent text to substitute in place of an inline tag, when Processing Tools create a visual preview for users. This attribute is optional, because the inline markup that the tag encapsulates may not be representable as text characters.

## <group>

In XLIFF:doc, the <group> element is used for a single purpose: to group <trans-unit> elements that can be joined without damaging the target document structure. Any sequential TUs within a <group> can be joined with each other. TUs from different groups may not be joined.

**Parent Elements:** <body>

**Child Elements:** <trans-unit>+

**Required Attributes:** *id*

#### *id*

The *id* must be unique within the <file>, but it ids can be recycled from file to file.

Acceptable values: any value that is unique within the <file> element.

**Optional Attributes:** none

## <header>

**Parent Elements:** <file>

**Child Elements:** <skl>{1,1}, <dx:note>\*, <tool>{1,1}, <dx:internal-preview> {1,1}, <dx:external-xsl>?

**Required Attributes:** none

**Optional Attributes:** none

## Interoperability

### <mrk>

This element has one function in XLIFF:doc: to provide a look-up reference to the embedded UTX glossary. This element is placed around a piece of text identified as being a term in the glossary. It can be used to demarcate terms in both the source and target.

**Parent Elements:** <source>, <target>

**Child Elements:** <g>\*, <x>\*

**Required Attributes:** *mtype*, *dx:term-id*

*mtype*

The value for this attribute in XLIFF:doc must be "term".

*dx:term-id*

Reference to the *entry-id* attribute of the <dx:utx-term> that contains a term designated as a match or hit for this element's grand-parent <trans-unit>.

**Optional Attributes:** none

### <skl>

Use <external-file> to point to the original skeleton file (structural content of the original source file). XLIFF:doc does not support <internal-file> (embedded structure content within the XLIFF).

**Parent Elements:** <header>

**Child Elements:** <external-file>{1,1}

**Required Attributes:** none

**Optional Attributes:** none

### <source>

**Parent Elements:** <trans-unit>, <alt-trans>

**Child Elements:** <g>\*, <x>\*, <mrk>\*

**Required Attributes:** none

**Optional Attributes:** none

The contents of the <source> element are considered immutable, with one exception. When a processing tool merges two or more TUs, it will create a new <source> element that contains the contents of the original <source> elements. In all other situations, the contents of the <source> element are not to be edited, deleted, or adjusted in any other way.

### <target>

**Parent Elements:** <trans-unit>, <alt-trans>

**Child Elements:** <g>\*, <x>\*, <mrk>\*

**Required Attributes:** *state*

*state*

Acceptable values: "new", "translated", "final", "signed-off", "needs-review-translation". Default value: "new".

**Optional Attributes:** none

## Interoperability

### <tool>

See the XLIFF 1.2 specification for more information. In XLIFF:doc, all 4 attributes are required, and must refer to the name of the tool that originally created the XLIFF. The values must not be modified by Processing Tools (to refer to themselves).



Example:

```
<tool tool-id="1" tool-name="MyTMserver" tool-version="v2.0.1" tool-company="Acme Rocketry Corporation" />
```

**Parent Elements:** <header>

**Child Elements:** none

**Required Attributes:** *tool-id, tool-name, tool-version, tool-company*

**Optional Attributes:** none

### <trans-unit>

In XLIFF:doc, the TU can be assigned information about the match that was applied to it (if any). This information must be assigned to a TU if a match is applied from any source: embedded <alt-trans> matches, live TM lookup, live MT application, etc. If, however, subsequent to applying a match, the user edits the <target> element, then the match-related attributes (*dx:match-quality, dx:glorious-match, dx:match-penalty, origin*) must be removed, as the translation no longer represents the match as applied.

**Parent Elements:** <group>

**Child Elements:** <source>{1,1}, <target>{1,1}, <alt-trans>\*, <dx:note>\*, <dx:qa-hits>\*

**Required Attributes:** *id, dx:match-quality, dx:glorious-match, dx:repetition, dx:match-penalty, dx:origin, dx:origin-shorttext, dx:modified-by, dx:modified-at*

*id*

Acceptable values: any value that is unique within the XLIFF:doc file.

*dx:match-quality*

Use values 0-100. The use of decimal numbers is acceptable, but only a single digit of precision is allowed. For example, "99.5" is acceptable, but "99.52" is not. Do not use a percent sign. Higher numbers refer to better matches than lower numbers.

Remove this attribute if the user manually edits the <target> after the match was placed.



*Note: this mimics the match-quality attribute of <alt-trans> in XLIFF 1.2, giving us the same functionality for the <trans-unit> element.*

*dx:glorious-match*

Indicates whether the match is an exact, in-context match (as determined by the tool that generated the match). Remove this attribute if the user manually edits the <target> after the match was placed. Acceptable values: "yes", "no". Default value: "no".

*dx:repetition*

Indicates whether this translation unit has been classified as a repetition (same exact source as other translations in the same project). XLIFF:doc does not specify how to determine if a TU is a repetition or not (2 occurrences vs 3, etc.). XLIFF:doc does not distinguish between first instance of a repetition and subsequent instances. Acceptable values: "yes", "no". Default value: "no".

## Interoperability

### *dx:origin*

Acceptable values: "TM", "MT", "XLIFF", "manual", "sourcetext", "pseudo-trans". If the match was pulled from the `<alt-trans>` in the XLIFF:doc doc, this value must match the *origin* value of the `<alt-trans>` match.

Value	Description
TM	Indicates the match is from translation memory
MT	Match is from machine translation of some type
XLIFF	Match was copied directly from another XLIFF file
manual	A user edited the <code>&lt;target&gt;</code> element directly (by typing, pasting text, etc.)
sourcetext	The source was copied to the target via some automated means (a pre-translation task, or user selected Edit > Copy Source to Target, etc.). If a user copies the source text themselves, and pastes into the target, that would be "manual".
pseudo-trans	Some type of pseudo-translation function has populated the target. For example, an "X" was placed in the target for each character in the source text.

☀ *Note: this mimics the *origin* attribute of XLIFF 1.2, and allows it to be used in the `<trans-unit>` element.*

### *dx:origin-shorttext*

Human-readable string describing the match origin. This could be the name of the TM it came out of, the name of an MT engine, etc. Examples: "Master TM 1", "Google MT", "myDocument.docx.xlf", "Manual Translation". Exact values are not defined by XLIFF:doc.

### *dx:match-penalty*

Identifies what, if any, artificial (non-textual) penalty was applied to the match by the tool that generated the match. Use values 0-100. The use of decimal numbers is acceptable, but only a single digit of precision is allowed. For example, "0.5" is acceptable, but "0.48" is not. Do not use a percent sign. Higher numbers refer to larger penalties than lower numbers.

Remove this attribute if the user manually edits the `<target>` after the match was placed.

### *dx:modified-by*

String value identifying the user (human or automatic) responsible for the most recent modification. Modification could include changing an attribute of the `<trans-unit>` element, or any change to the `<target>` element.

### *dx:modified-at*

Time and date the `<trans-unit>` was last modified. This can represent the date/time the translation was edited, or when a TM match was applied, or when the `<trans-unit>` was processed in some other way.

☀ *Note: Date/time must be specified in ISO 8601 format. Time must be specified in UTC (Zulu) time zone. [See the XLIFF date element for more information.](#)*

**Optional Attributes:** *translate*, *dx:xsl-hook*

### *translate*

Acceptable values: "yes", "no". If set to "no", this TU will be considered "locked", and shall not be editable by tool users. Default value: "yes".

## Interoperability

- ☼ *Note: Processing Tools are required to make all TUs in an XLIFF:doc document available to users, even locked ones. A Processing Tool may optionally provide a mechanism for users to intentionally hide locked TUs, but that decision must be available to user, and not made by the tool. For example, a Processing Tool that did not "import" locked TUs would not be XLIFF:doc compliant.*
- ☼ *Note: A Processing Tool must prevent users from editing the content of the <target> element, for any locked TU. However, XLIFF:doc does not require that a Processing Tool prevent a user from adding a note to a locked TU. Similarly, a Processing Tool may add QA hits to a locked TU.*

### dx:xsl-hook

This attribute is used to optionally provide information to an XSLT file, about the translation unit. This information would typically be used to help determine how the contents of the translation unit should be displayed in a preview. XLIFF:doc does not specify how this attribute should be populated or used.

Acceptable values: undefined. Must be a string value.

### <x>

**Parent Elements:** <source>, <target>

**Child Elements:** none

**Required Attributes:** *id*, *dx:orig-markup*, *dx:equiv-markup*, *ctype*

### *id*

The *id* is the primary mechanism for matching placeholders from source to target. The *id* must be unique within the <trans-unit>, but it ids can be recycled from trans-unit to trans-unit. An *id* used for a <g> element cannot be re-used for an <x> element within the same <trans-unit>.

Acceptable values: non-negative integers

### *dx:orig-markup*

If the original markup can be escaped safely and included, put it into this attribute. If the original markup cannot be safely encoded in XML, do not include this attribute.

### *dx:equiv-markup*

HTML code that approximates the formatting of the original document format. For example, if the original markup was a line break, the *dx:equiv-markup* might be "<br>".

- ☼ *Note: if the original format is HTML, and the HTML formatting can be represented safely, it is not necessary to include dx:equiv-markup attributes. If not present, the dx:orig-markup will be used.*

### *ctype*

XLIFF:doc allows the following values for *ctype* in the <x> element:

Value	Description	Original HTML Markup Allowed?
image	Indicates a inline image.	Yes
pb	Indicates a page break.	No
lb	Indicates a line break.	No
x-xref	Indicates a cross-reference.	No

## Interoperability

x-tab	Indicates a tab character.	No
x-other	All other types.	No

### Optional Attributes: *equiv-text*

#### *equiv-text*

Indicates the equivalent text to substitute in place of an inline tag, when Processing Tools create a visual preview for users. This attribute is optional, because the inline markup that the tag encapsulates may not be representable as text characters.

### <xliff>

**Parent Elements:** none

**Child Elements:** <file>+, <dx:utx-glossary>\*

**Required Attributes:** *dx:version*, *version*, *xmlns*, *xmlns:dx*, *xmlns:xsi*, *xsi:schemaLocation*

In XLIFF:doc, the <xliff> element contains critical information which must be set in order for an XLIFF:doc to be recognized as an XLIFF:doc file, and also to allow for proper validation of the XML. With the exception of the *dx:version* attribute, all attributes and attribute values are fixed.



Example:

```
<xliff version="1.2" xmlns="urn:oasis:names:tc:xliff:document:1.2" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:dx="http://www.interoperability-now.org/schema" xsi:schemaLocation="urn:oasis:names:tc:xliff:document:1.2 xliff-doc-1_0_extensions.xsd" dx:version="1.4">
```

#### *version*

The *version* attribute specifies the version of XLIFF. At the current time, the only allowed version is "1.2".

Acceptable values: "1.2"

#### *dx:version*

The *dx:version* attribute specifies the version of the XLIFF:doc XSD used.

Acceptable values: non-negative numbers corresponding to published XLIFF:doc XSD versions.

#### *xmlns*

The *xmlns* namespace declaration for XLIFF must be included.

Acceptable values: "urn:oasis:names:tc:xliff:document:1.2"

#### *xmlns:dx*

The *xmlns:dx* namespace declaration for XLIFF:doc must be included.

Acceptable values: "http://www.interoperability-now.org/schema"

#### *xmlns:xsi*

The *xmlns:xsi* namespace declaration is currently required.

Acceptable values: "<http://www.w3.org/2001/XMLSchema-instance>"

`xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"`

#### *xsi:schemaLocation*

## Interoperability

XLIFF:doc currently requires the following XSI callout, which helps manage the limitations XLIFF:doc places on the XLIFF 1.2 specification.

Acceptable values: "urn:oasis:names:tc:xliff:document:1.2 xliff-doc-1\_0\_extensions.xsd"




## Appendix C: Version History

Version	Date	Author	Change(s)
1.0.1	2013/04/25	Micah Bly	<ul style="list-style-type: none"><li>• No changes were made to schema, compared to 1.0 version of this document.</li><li>• Fixed a code example that was missing the dx: before a dx: attribute.</li><li>• Fixed a code example that showed the wrong child order for the <code>&lt;alt-trans&gt;</code> element (page 21).</li><li>• Updated the XLIFF:doc schema diagram to reflect required element order for the <code>&lt;trans-unit&gt;</code> and <code>&lt;alt-trans&gt;</code> elements' children.</li></ul>

Version	Date	Author	Change(s)
1.0	2013/03/01	Micah Bly	<ul style="list-style-type: none"> <li>Fixed specification of child elements for <code>&lt;g&gt;</code> and <code>&lt;mrk&gt;</code> in the reference section.</li> <li>References to languages now point to RFC 5646, which is the current successor to BCP 47.</li> <li>Switched from using the XLIFF 1.2 native <code>&lt;note&gt;</code> element to a new <code>&lt;dx:note&gt;</code> element. This change was required to maintain backwards compatibility with XLIFF 1.2, and still be able to assign a date-time to a note. In XLIFF 1.2, the <code>&lt;note&gt;</code> element cannot be extended with new attributes, so adding <code>dx:modified-at</code> to it made the resulting XLIFF:doc invalid.</li> <li>Removed <code>&lt;dx:external-preview&gt;</code> element and the option to have the HTML preview by external to the XLIFF:doc. The HTML preview must be embedded into the XLIFF:doc file itself, using the <code>&lt;dx:internal-preview&gt;</code> element.</li> <li>Replaced the placeholder notes about project meta data with text explaining that XLIFF:doc does not have a mechanism for project meta data, and that people interested in that may want to consult the TIPP format.</li> <li>Added text to the <code>&lt;dx:utx-glossary&gt;</code> element description explaining that only languages that match the XLIFF:doc's languages can be used in a glossary.</li> <li>Added text to the <code>&lt;trans-unit&gt;</code> element description detailing expected behavior around locked TUs (translate=no).</li> <li>Added text to <code>&lt;source&gt;</code> element description explicitly stating when the contents of a <code>&lt;source&gt;</code> element can be modified (only when merging TUs).</li> <li>Added a sub-section on whitespace between TUs to the Structure, Content, and Preview chapter, explicitly stating how spaces between segments are to be handled by Creation Tools and Processing Tools.</li> <li>Added ability to embed textual context for embedded TU matches: <ul style="list-style-type: none"> <li>Added description of functionality to the Translation Matches section, and modified <code>&lt;alt-trans&gt;</code> examples found in that section.</li> <li>Added write-up of new context elements: <code>&lt;dx:context&gt;</code>, <code>&lt;dx:context-source-before&gt;</code>, <code>&lt;dx:context-source-after&gt;</code>, <code>&lt;dx:context-target-before&gt;</code>, <code>&lt;dx:context-target-after&gt;</code></li> </ul> </li> <li>Tweaked the versioning functionality: version history is now a read-only set of information embedded, optionally, by the Creation Tool. The reason for limiting version history to Creation Tool-only is because XLIFF:doc files are temporary mechanisms for exchanging data between 2 arbitrary systems, and meant to be passed as a semi-permanent object, from tool to tool to tool without being re-created.</li> </ul>

## Interoperability

Version	Date	Author	Change(s)
0.953	2012/06/09	Micah Bly	<ul style="list-style-type: none"> <li>Changed <code>&lt;note&gt;</code> element: it no longer uses the merged "from" attribute to indicate both a user and a time/date. The time/date is now saved in a <code>dx:modified-at</code> attribute.</li> <li>Fixed errors in the <code>dx:allow-tb-match-addition</code> and <code>dx:allow-tm-match-addition</code> descriptions: said "yes" where they needed to say "no". Thanks to Jörg Schütz for reporting the error.</li> </ul>
0.952	2012/03/20	Micah Bly	<ul style="list-style-type: none"> <li>Clarified that the date-created attribute in the UTX glossary is a timestamp (datetime), not a date (only) value.</li> <li>Corrected entries for <code>dx:utx-related</code> items to reflect fact that the attributes used need to be prefaced by dx:, as they are part of the dx: namespace, not the XLIFF 1.2 namespace.</li> </ul>
0.951	2012/03/12	Micah Bly	<ul style="list-style-type: none"> <li>Corrected list of acceptable values for <code>ctype</code> for <code>&lt;g&gt;</code> and <code>&lt;x&gt;</code>, to make all values compliant with XLIFF 1.2.</li> </ul>
0.95	2012/01/03	Micah Bly	<ul style="list-style-type: none"> <li>Removed the <code>&lt;dx:term-hit&gt;</code> and <code>&lt;dx:term-hits&gt;</code> elements as the method of referencing embedded terms. Terms are now referenced via the <code>&lt;mrk&gt;</code> element and its <code>dx:term-id</code> attribute.</li> </ul>
0.941	2011/12/19	Micah Bly	<ul style="list-style-type: none"> <li>Added this version history table</li> <li>Removed "[TEMP] Misc. Questions" Appendix.</li> <li>clarified usage of the datatype attribute, and hardened the pattern for acceptable values.</li> <li>removed "x-" extensions in front of ctype values. The values are now fixed, and no extensions are allowed. There are "other" and "other formatting" values available to cover n-cases.</li> <li>added specification text for the equiv-text attributes in the reference section.</li> </ul>
			 <i>Note: See SVN log for previous version history</i>