

Review N: Translation and localization

Table of contents

1 Translation and localization.....	3
1.1 The @xml:lang attribute.....	3
1.1.1 Recommendations for the @xml:lang attribute.....	3
1.1.2 Processing expectations regarding the @xml:lang attribute.....	4
1.1.3 Example: content reference and the @xml:lang attribute.....	4
1.2 The @dir attribute.....	5
1.2.1 The Unicode Bidirectional Algorithm.....	5
1.2.2 Recommended usage of the dir attribute.....	6
1.2.3 Processing expectations regarding the Unicode Bidirectional Algorithm.....	6
1.3 The @translate attribute.....	6
A Aggregated RFC-2119 statements.....	7
Index.....	8

1 Translation and localization

DITA has markup that facilitates translation and localization. This markup includes the `@xml:lang` attribute, the `@dir` attribute, and the `@translate` attribute.

1.1 The `@xml:lang` attribute

The `@xml:lang` attribute specifies the language and optional locale of the content that is contained in an element. The `@xml:lang` attribute is described in the [XML Recommendation](#).

Since the `@xml:lang` attribute is an inherent property of the XML document, it does not behave in the same way as other DITA metadata attributes do.

Within topic and map documents, the `@xml:lang` attribute applies to the content and attributes that are contained by the element on which it is specified. This means that it supplies a value for lower-level elements in the containment hierarchy that do not supply their own value for the `@xml:lang` attribute. However, any such value is overridden when an `@xml:lang` attribute with a different value is specified on lower-level elements in the containment hierarchy.

When the `@xml:lang` attribute is specified on a topic reference, it does not apply to the referenced resource. This means that the value of the `@xml:lang` attribute on a topic reference (or the root element of the map) does not automatically supply a default value for the referenced topic or DITA map.

For topic and map documents, if no value for the `@xml:lang` value is specified explicitly or on a higher-level element in the containment hierarchy, a processor-determined default value is assumed.

1.1.1 Recommendations for the `@xml:lang` attribute

Specifying the `@xml:lang` attribute in the DITA source facilitates translation and helps ensure that processors will handle content appropriately. Accordingly, this specification makes certain recommendations for best practices for where the `@xml:lang` attribute should be set.

Setting the `@xml:lang` attribute in the source-language document facilitates the translation process. Some translation tools do not support adding new markup to the document that is being translated, so if the source language content does not set the `@xml:lang` attribute, it might be difficult or impossible for a translator to add the `@xml:lang` attribute to the translated document.

In addition, setting the `@xml:lang` attribute in the DITA source ensures that processors handle content in a language- and locale-appropriate way. If the `@xml:lang` attribute is not set, processors might assume a default value which is inappropriate for the DITA content.

The following table outlines the recommended use of the `@xml:lang` attribute in topics and maps. These recommendations ensure that DITA resources have an effective default language.

DITA resource	Recommended use
DITA topic document that contains a single language	Specify the <code>@xml:lang</code> attribute on the root element of the document.
DITA topic document that contains more than one language	Specify the primary language and locale that applies to the topic on the highest-level element that contains content. If part of a topic is written in a different language, enclose that content in an element with the <code>@xml:lang</code> attribute set appropriately. This applies to

DITA resource	Recommended use
	both block and inline elements that use the alternate language.
DITA map	Specify the @xml:lang attribute on the root element of the map. This applies both to the root map and any submaps.

1.1.2 Processing expectations regarding the @xml:lang attribute

When the @xml:lang attribute is specified as recommended, a language for the content is clearly indicated. However, when the @xml:lang attribute is not specified, processors might need to assign a default value.

- 001 (7) If the root element of a map or a top-level topic has no value for the @xml:lang attribute, a processor **SHOULD** assume a default value. The default value of the processor can be either fixed, configurable, or derived from the content itself, such as the @xml:lang attribute on the root map.
- 002 (7) When a @conref or @conkeyref attribute is used to include content from one element into another, the processor **MUST** use the effective value of the @xml:lang attribute from the referenced element.
- 003 (7) Processors **SHOULD** render each element in a way that is appropriate for its language as identified by the @xml:lang attribute.

1.1.3 Example: content reference and the @xml:lang attribute

This scenario outlines how processors determine the effective value of the @xml:lang attribute for content that is referenced by the @conref or @conkeyref attribute.

Comment by Kristen J Eberlein on 28 March 2022

This scenario kind of exemplifies an edge case of an edge case ... I don't know how many examples of @xml:lang we need, but this one seems very contrived.

In this example, a DITA topic specifies an @xml:lang attribute of "en". However, it contains content that is conrefed from a element that explicitly sets the @xml:lang attribute to "fr".

The following code block shows the content of the DITA topic that contains the referencing element:

```
<?xml version="1.0"?>
<!DOCTYPE task PUBLIC "-//OASIS//DTD DITA Task//EN" "task.dtd">
<task xml:lang="en" id="install_acme">
  <title>Installing Acme</title>
  <shortdesc>Step-by-step details about how to install the Acme product.</shortdesc>
  <taskbody>
    <prereq>
      <p>Special notes when installing Acme in France:</p>
      <note id="mynote" conref="warningsAcme.dita#warnings/french1"/>
    </prereq>
  </taskbody>
</task>
```

The following code block shows the content of the topic that contains the referenced element:

```
<?xml version="1.0"?>
<!DOCTYPE topic PUBLIC "-//OASIS//DTD DITA Topic//EN" "topic.dtd">
```

```

<topic id="warnings">
  <title>Warnings</title>
  <body>
    <section xml:lang="fr">
      <title>French warnings</title>
      <p>These are our French warnings.</p>
      <note id="french1">Note in French!</note>
      <!-- ... -->
    </section>
    <section xml:lang="en">
      <title>English warnings</title>
      <p>These are our English warnings.</p>
      <note id="english1">Note in English!</note>
      <!-- ... -->
    </section>
  </body>
</topic>

```

Note that the value of the `@xml:lang` attribute of the referenced note is set by the containing `<section>` element that sets the `@xml:lang` attribute to "fr". When the topic that contained the referenced note is processed, the `<note>` element with the `@id` attribute set to "mynote" has an effective value for the `@xml:lang` attribute of "fr".

1.2 The `@dir` attribute

The `@dir` attribute provides instructions to processors about how bidirectional text is rendered.

The `@dir` attribute identifies or overrides the text directionality. The following values are valid:

lro

Indicates an override of the Unicode Bidirectional Algorithm, forcing the element into left-to-right mode.

ltr

Indicates left-to-right.

rlo

Indicates an override of the Unicode Bidirectional Algorithm, forcing the element into right-to-left mode.

rtl

Indicates right-to-left.

-dita-use-conref-target

See for more information.

1.2.1 The Unicode Bidirectional Algorithm

The Unicode Bidirectional Algorithm plays a critical role in ensuring that bidirectional text is correctly rendered.

Bidirectional text is text that contains text in both text directionalities, right-to-left (RTL) and left-to-right (LTR). Common examples of bidirectional text include the following:

- Documents in RTL languages such as Arabic, Hebrew, Farsi, Urdu, and Yiddish that include numerics or embedded sections of LTR text
- Documents that contain text in both LTR and RLT languages, for example, a topic that **lists** the names of a **movie** in multiple languages

The Unicode Bidirectional Algorithm specifies how text should be rendered for a given language. For more information about the Unicode Bidirectional Algorithm, see the following resources:

- [Unicode Bidirectional Algorithm, Unicode Standard Annex #9](#)

- [Specifying the direction of text and tables: the dir attribute](#), HTML 4.01 Specification
- [Inline markup and bidirectional text in HTML](#), W3C internationalization article
- [XHTML Bi-directional Text Attribute Module](#), XHTML 2.0 W3C Working Draft 22

1.2.2 Recommended usage of the dir attribute

Typically, processors that fully support the Unicode Bidirectional Algorithm handle bidirectional text without the need to specify directionality in the DITA source, if the `@xml:lang` attribute is specified on the highest-level element.

The need to specify the `@dir` attribute primarily occurs in the following situations:

- Processors that do not fully support the Unicode Bidirectional Algorithm
- Documents that contain bidirectional text and characters with neutral bidirectionality

For the above situations, we recommend that DITA source documents, in addition to specifying the `@xml:lang` attribute, also specify the `@dir` attribute on the highest-level element that is necessary.

1.2.3 Processing expectations regarding the Unicode Bidirectional Algorithm

Processor support for the Unicode Bidirectional Algorithm is critical.

004 (7) DITA processors **SHOULD** fully support the Unicode Bidirectional Algorithm. This ensures that processors can implement the script and directionality for each language that is used in a document.

1.3 The @translate attribute

The `@translate` attribute provides information about whether the content of an element is to be translated.

The following values are valid: "yes", "no", and "-dita-use-conref-target".

A few elements have the `@translate` attribute set by default to "no". These elements include `<draft-comment>` and `<required-cleanup>`, all elements that are designed to hold content that is not intended for publication.

The non-normative appendix, [Element-by-element recommendations for translators](#), includes information on whether the element is block or inline, whether the element contents are likely to be suitable for translation, and whether the element has attributes whose values might need translation.

A Aggregated RFC-2119 statements

This appendix contains all the normative statements from the DITA 2.0 specification. They are aggregated here for convenience in this non-normative appendix.

Item	Conformance statement
001 (4)	If the root element of a map or a top-level topic has no value for the <code>@xml:lang</code> attribute, a processor SHOULD assume a default value. The default value of the processor can be either fixed, configurable, or derived from the content itself, such as the <code>@xml:lang</code> attribute on the root map.
002 (4)	When a <code>@conref</code> or <code>@conkeyref</code> attribute is used to include content from one element into another, the processor MUST use the effective value of the <code>@xml:lang</code> attribute from the referenced element.
003 (4)	Processors SHOULD render each element in a way that is appropriate for its language as identified by the <code>@xml:lang</code> attribute.
004 (6)	DITA processors SHOULD fully support the Unicode Bidirectional Algorithm. This ensures that processors can implement the script and directionality for each language that is used in a document.

Index

B

bi-directional text [5](#)

D

@dir attribute [5](#)

P

processing expectations
bi-directional text [5](#)