
**Language resource management —
Lexical markup framework (LMF) —**

**Part 4:
TEI serialization**

*Gestion des ressources linguistiques — Cadre de balisage lexical
(LMF) —*

Partie 4: Sérialisation TEI

STANDARDSISO.COM : Click to view the full PDF of ISO 24613-4:2021



STANDARDSISO.COM : Click to view the full PDF of ISO 24613-4:2021



COPYRIGHT PROTECTED DOCUMENT

© ISO 2021

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword	v
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
4 General principles	1
5 Serialization of the LMF core model (ISO 24613-1)	2
5.1 Implementing the LexicalResource class	2
5.2 Implementing the GlobalInformation class	2
5.3 Implementing the Lexicon class	3
5.4 Implementing the LexiconInformation class	4
5.5 Implementing the LexicalEntry class	4
5.6 Implementing the Form and Lemma classes	5
5.6.1 Form class	5
5.6.2 Lemma class	5
5.7 Implementing the GrammaticalInformation class	5
5.8 Implementing the Sense class	7
5.9 Implementing the Definition class	7
5.10 Implementing the OrthographicRepresentation class	8
5.11 Implementing the CrossREF class	8
5.12 Data category selection	9
6 Serialization of the MRD model (ISO 24613-2)	9
6.1 Implementing the form representations for the Form class	9
6.2 Implementing classes derived from the Form class	11
6.2.1 General principles	11
6.2.2 Implementing the WordForm class	11
6.2.3 Implementing the Stem class	11
6.2.4 Implementing the WordPart class	11
6.2.5 Implementing the RelatedForm class	12
6.3 Implementing the Bibliography class	12
6.4 Implementing the CrossREF mechanism to refer to external media files	12
6.5 Implementing the Example class	12
6.6 Implementing the Translation class	13
6.7 Implementing the TextRepresentation class	13
6.8 Implementing the SubjectField class	14
6.9 Implementing the CrossREF mechanism to represent related entries	14
7 Implementing the classes from the etymological extension (ISO 24613-3)	15
7.1 Implementing the Etymology class	15
7.2 Implementing the Etymon class	15
7.2.1 Referencing forms in an etymon	16
7.2.2 Representing the meaning of an etymon	16
7.2.3 Representing the language of an etymon	16
7.2.4 Associating grammatical information to an etymon	16
7.2.5 Dating an etymon	16
7.2.6 Providing sources associated with an etymon	17
7.2.7 Further prose component in an etymological description	17
7.3 Implementing the EtyLink class	17
7.4 Implementing the CognateSet class	17
7.5 Implementing the Cognate class	17
7.6 Implementing the Gloss class	18
8 Additional mechanisms	18
8.1 Overview	18
8.2 Representing punctuation marks with <pc>	18

8.3	Representing various labels with <lbl>	18
8.4	Encoding simple separating characters with <metamark>	19
8.5	Providing rendition information with the @rend attribute.....	19
8.6	Usage example of the additional features introduced in Clause 8	19
Bibliography		20

STANDARDSISO.COM : Click to view the full PDF of ISO 24613-4:2021

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 37, *Language and terminology*, Subcommittee SC 4, *Language resource management*.

This first edition of ISO 24613-4, together with ISO 24613-1:2019, ISO 24613-2:2020, ISO 24613-3:—¹⁾ and ISO 24613-5:—²⁾, cancels and replaces ISO 24613:2008, which has been technically revised.

The main changes compared to the previous edition are as follows:

- entire revision of the content and its subdivision into several parts.

A list of all parts in the ISO 24613 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

1) Under preparation. Stage at the time of publication: ISO/FDIS 24613-3:2020.

2) Under preparation. Stage at the time of publication: ISO/DIS 24613-5:2020.

STANDARDSISO.COM : Click to view the full PDF of ISO 24613-4:2021

Language resource management — Lexical markup framework (LMF) —

Part 4: TEI serialization

1 Scope

This document describes the serialization of the lexical markup framework (LMF) model defined as an XML model compliant with the Text Encoding Initiative (TEI) Guidelines. This serialization covers the classes of ISO 24613-1 (the LMF core model) as well as classes provided by ISO 24613-2 (the machine readable dictionary, MRD, model) and ISO 24613-3 (the etymological extension).

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 24613-1, *Language resource management — Lexical markup framework (LMF) — Part 1: Core model*

ISO 24613-2, *Language resource management — Lexical markup framework (LMF) — Part 2: Machine-readable dictionary (MRD) model*

ISO 24613-3, *Language resource management — Lexical markup framework (LMF) — Part 3: Etymological extension*

IETF BCP 47, *Tags for Identifying Languages*. (ed A. Phillips, M. Davis). September 2009. Best Current Practice. <https://tools.ietf.org/html/bcp47>

W3C XML Recommendation, *Extensible Markup Language (XML) 1.0* (Fifth Edition), 26 November 2008, <http://www.w3.org/TR/xml/>

TEI P5, *Guidelines for Electronic Text Encoding and Interchange*. [Version number: 4.1.0]. [Last modified date: 2020-08-19] TEI Consortium. <http://www.tei-c.org/Guidelines/P5/>

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 24613-1 and in ISO 24613-3 apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

- ISO Online browsing platform: available at <https://www.iso.org/obp>
- IEC Electropedia: available at <http://www.electropedia.org/>

4 General principles

This document aims at providing mostly univocal constructs for each LMF class from the core model (ISO 24613-1), the MRD extension (ISO 24613-2) and the etymological extension (ISO 24613-3), thus making the original Text Encoding Initiative (TEI) recommendations from the dictionary chapter

terser (see References [4], [8], [9], [11] and [14]), while eliciting specific constraints needed to align with ISO 24613 in general. For instance, precise value lists are given for some attributes such as @type. In addition, this document complies with the cardinalities expressed in ISO 24613-1, ISO 24613-2 and in ISO 24613-3.

Unless explicitly stated, all resulting constructs shall be valid TEI representations, which means that the specification described in this document corresponds to a pure subset of the TEI Guidelines. They shall thus be well-formed XML documents as specified by the W3C XML recommendation.

This document requires compliance with ISO 24613-1, ISO 24613-2, and ISO 24613-3 when implementing data categories referred to in the respective parts.

Nevertheless, this document does not elaborate on the metadata aspects from LMF, since the TEI header, i.e. the metadata component attached to any TEI document, is in essence rich enough in that it represents all the aspects related to the creation, the content description, the versioning and publishing of a textual document as a whole.

In all XML examples in this document and in order to simplify the actual representations, it is assumed, unless otherwise stated, that XML elements belong to the TEI namespace, thus assuming that all examples are within the scope of the following XML namespace declaration:

```
xmlns="http://www.tei-c.org/ns/1.0"
```

An updated TEI ODD specification corresponding to this document is available from Reference [12] together with additional examples.

5 Serialization of the LMF core model (ISO 24613-1)

5.1 Implementing the LexicalResource class

The LexicalResource class shall be implemented by means of the <teiCorpus>³⁾ element which groups together one or several TEI documents for each specific lexicon. The <teiCorpus> element shall include a @type attribute with the value "lexicalResource" (see Table 1). In the case where the lexical resource contains only one lexicon, this level may be omitted, so that the document directly starts with a <TEI> root element (see 5.3).

Table 1 — LexicalResource class

LMF class	TEI construct
/LexicalResource/	<teiCorpus type="lexicalResource"> ... </teiCorpus>

5.2 Implementing the GlobalInformation class

The GlobalInformation class shall be implemented by means of the TEI element <teiHeader>⁴⁾, as a direct child of the <teiCorpus> element (see Table 2). For the actual use of the <teiHeader> element, the additional constraints given in the TEI Guidelines shall be followed when encoding any kind of information related to administrative, documentary and bibliographic information attached to the corresponding lexical resource.

Table 2 — GlobalInformation class

LMF class	TEI construct
/GlobalInformation/	<teiHeader> ... </teiHeader>

3) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-teiCorpus.html>

4) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-teiHeader.html>

The further requirements concerning the GlobalInformation class are implemented as follows:

- since the TEI serialization is based on the W3C XML recommendation, it implements the @xml:lang attribute to indicate the working language information corresponding to the content of any specific element. @xml:lang content shall be compliant with IETF BCP 47. There is thus no need for a specific implementation of the /language coding/ data category in order to ensure compliance of this document with ISO 24613-1;
- IETF BCP 47 also relies on ISO 15924 for the coding of script names. There is thus no need to have a specific mechanism in this document for implementing the /script coding/ data category;
- the /character encoding/ data category is implemented in the XML declaration of a TEI conformant document using the *encoding* attribute. For instance, an XML-TEI document encoded as UTF-8 according to the Unicode standard ISO/IEC 10646 shall begin with the following declaration:

```
<?xml version="1.0" encoding="UTF-8" ?>
```

5.3 Implementing the Lexicon class

The Lexicon class shall be implemented by means of the TEI element <TEI>⁵⁾, which groups together all lexical entries of a specific lexicon. The <TEI> element shall have a @type attribute with the value "lexicon" (see Table 3).

Table 3 — Lexicon class

LMF class	TEI construct
/Lexicon/	<TEI type="lexicon"> ... </TEI>

According to the TEI Guidelines, the actual content of a TEI document is embedded within a <text>⁶⁾ child element which is further decomposed in sub-elements <front>⁷⁾, <body>⁸⁾, and <back>⁹⁾. All lexical entries of a given lexicon shall be descendant elements of the <body> sub-element of <text>, and may be further organized in subsets by means of the <div>¹⁰⁾ element. <front> and <back> may be used to incorporate complementary resources when applicable, for instance bibliographic references.

The following example outlines a simple document organization for a lexicon (without the optional <front> and <back> elements, and without sub-divisions).

EXAMPLE

```
<TEI type="lexicon">
  <teiHeader> ... </teiHeader>
  <text>
    <body>
      <entry> ... </entry>
      <entry> ... </entry>
      ...
    </body>
  </text>
</TEI>
```

5) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-TEI.html>

6) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-text.html>

7) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-front.html>

8) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-body.html>

9) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-back.html>

10) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-div.html>

5.4 Implementing the LexiconInformation class

The LexiconInformation class shall also be implemented by means of the TEI element <teiHeader>, as a direct child of the <TEI> element (see Table 4). For the actual use of the <teiHeader> element, the additional constraints given in the TEI Guidelines shall be followed to encode any kind of information related to administrative, documentary and bibliographic information attached to the corresponding lexicon.

Table 4 — LexiconInformation class

LMF class	TEI construct
/LexiconInformation/	<teiHeader> ... </teiHeader>

5.5 Implementing the LexicalEntry class

The LexicalEntry class shall be implemented by means of the <entry>¹¹⁾ element in the TEI Guidelines (see Table 5). Lexical information inside <entry> elements shall be encoded through the use of the following child elements:

- <form>¹²⁾ for morphological and grammatical information related to the whole entry (see 5.6.1 and 5.6.2);
- <sense>¹³⁾ for semantic information, see 5.8;
- <etym>¹⁴⁾ for etymological aspects, see Clause 7;
- <re>¹⁵⁾ related entry allows the serialization of information referring to entry-like constructs (e.g., related forms, derived forms, multiword expressions), see 6.9;
- <xr>¹⁶⁾ for referencing internal or external elements (phrase, equivalent, etc.);
- <dictScrap>¹⁷⁾ for any information that does not belong to any of the above elements.

These various constructs are further described in the following subclauses.

Table 5 — LexicalEntry class

LMF class	TEI construct
/LexicalEntry/	<entry> ... </entry>

The following example in French illustrates the encoding of a simple dictionary entry¹⁸⁾.

EXAMPLE

```
<entry xml:lang="fr">
  <form type="lemma">
    <orth>langouste</orth>
    <pron>lɑ̃gust</pron>
    <gramGrp>
      <pos>n.</pos>
    </gramGrp>
  </form>
</entry>
```

11) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html>

12) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-form.html>

13) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-sense.html>

14) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-etym.html>

15) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-re.html>

16) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-xr.html>

17) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-dictScrap.html>

18) Source: Reference [10].

```

    <gen>f.</gen>
  </gramGrp>
</form>
  <sense n="1">
    <def>Grand crustacé marin (Décapodes macroures) aux pattes antérieures dépourvues
de pinces, aux antennes longues et fortes, et dont la chair est très appréciée.</def>
  </sense>
  <sense n="2">
    <usg type="socioCultural">Fig. et fam. (vulg.)</usg>
    <def>Femme, maîtresse.</def>
  </sense>
  <etym>XIIIe; languste, v. 1120, «sauterelle»; encore dans Corneille (Hymnes, 7); anc.
provençal langosta, altér. du lat. class. locusta «sauterelle».</etym>
</entry>

```

NOTE 1 The <usg> element is introduced later in this document.

NOTE 2 The etymological description is not completely encoded here. See [Clause 7](#).

5.6 Implementing the Form and Lemma classes

5.6.1 Form class

The Form class shall be serialized using the TEI element <form> (see [Table 6](#)). Further subclasses derived from the Form class are implemented by means of specific values for the @type attribute on <form>. See [Clause 6](#).

Table 6 — Form class

LMF class	TEI construct
/Form/	<form> ... </form>

5.6.2 Lemma class

The Lemma class is derived from the Form class, it shall thus be implemented as the TEI element <form> further constrained by means of the @type attribute set to “lemma” (see [Table 7](#)).

Table 7 — Lemma class

LMF class	TEI construct
/Lemma/	<form type="lemma"> ... </form>

5.7 Implementing the GrammaticalInformation class

The GrammaticalInformation class gathers grammatical features attached to the LexicalEntry, a Form, a Translation or a Sense (in the case of specific grammatical restrictions for instance). It shall be implemented in the TEI by means of the <gramGrp>¹⁹⁾ element (see [Table 8](#)) combined with various possible child elements for specific grammatical features (see below).

Table 8 — GrammaticalInformation class

LMF class	TEI construct
/GrammaticalInformation/	<gramGrp> ... </gramGrp>

19) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gramGrp.html>

The elements provided by the TEI Guidelines within <gramGrp> for grammatical description of an associated class are as follows.

- <pos>²⁰⁾ (part of speech) to indicate the grammatical category of the lexical item. This corresponds to the /partOfSpeech/ data category in ISO 24611:2012, Annex A.

NOTE 1 This element is equivalent to the use of the <gram> element with the appropriate @type attribute: <gram type="partOfSpeech">, as used in the TEI Lex 0 initiative^[5].

- <gen>²¹⁾ (grammatical gender) to indicate the grammatical gender (if relevant) of the lexical item or one of its inflected forms. This corresponds to the /grammaticalGender/ data category in ISO 24611:2012, Annex A.

NOTE 2 This element is equivalent to the use of the <gram> element with the appropriate @type attribute: <gram type="grammaticalGender">.

- <number>²²⁾ (grammatical number) to indicate the grammatical number (if relevant) of the lexical item or one of its inflected forms. This corresponds to the /grammaticalNumber/ data category in ISO 24611:2012, Annex A.

NOTE 3 This element is equivalent to the use of the <gram> element with the appropriate @type attribute: <gram type="grammaticalNumber">.

- <per>²³⁾ (person) to indicate the grammatical person (if relevant) of the lexical item or one of its inflected forms. This corresponds to the /person/ data category in ISO 24611:2012, Annex A.

NOTE 4 This element is equivalent to the use of the <gram> element with the appropriate @type attribute: <gram type="person">.

- <tns>²⁴⁾ (tense) to indicate the grammatical tense (if relevant) of the lexical item or one of its inflected forms. This corresponds to the /grammaticalTense/ data category in ISO 24611:2012, Annex A.

NOTE 5 This element is equivalent to the use of the <gram> element with the appropriate @type attribute: <gram type="grammaticalTense">.

- <subc>²⁵⁾ (subcategorization) to indicate subcategorization information (e.g. transitive/intransitive/ditransitive, countable/non-countable, etc.).

- <iType>²⁶⁾ (inflectional class) to indicate the inflectional class associated with the lexical item.

For instance, to indicate that the part of speech is a verb and that it is intransitive, the following construct shall be used, see Examples 1 to 3.

EXAMPLE 1

```
<gramGrp>
  <pos>verb</pos>
  <subc>intransitive</subc>
</gramGrp>
```

20) <http://tei-c.org/release/doc/tei-p5-doc/en/html/ref-pos.html>

21) <http://tei-c.org/release/doc/tei-p5-doc/en/html/ref-gen.html>

22) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-number.html>

23) <http://tei-c.org/release/doc/tei-p5-doc/en/html/ref-per.html>

24) <http://tei-c.org/release/doc/tei-p5-doc/en/html/ref-tns.html>

25) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-subc.html>

26) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-iType.html>

EXAMPLE 2 with a simple indication of subcategorization information:

```
<entry xml:lang="fr">
  <form type="lemma">
    <orth>pacifier</orth>
    <pron>pasifje</pron>
    <gramGrp>
      <pos>v.</pos>
      <subc>t.</subc>
    </gramGrp>
  </form>
  ...
</entry>
```

EXAMPLE 3 comprising the indication of a root together with its inflectional scheme in Arabic:

```
<entry xml:id="Kataba" xml:lang="ar">
  <form type="lemma">
    <orth>كتب</orth>
    <gramGrp>
      <pos>verb</pos>
    </gramGrp>
  </form>
  <form type="root">
    <orth>كتب</orth>
    <gramGrp>
      <pos norm="root">جذر</pos>
      <iType type="scheme">فعل</iType>
    </gramGrp>
  </form>
</entry>
```

Further recommendations for best TEI serialization practices of grammatical information are highlighted in Reference [5].

5.8 Implementing the Sense class

The Sense class, as a recursive construct, is implemented by means of the <sense> element in the TEI Guidelines (see Table 9). When senses have to be numbered, an @n attribute can be attached to this element.

Table 9 — Sense class

LMF class	TEI construct
/Sense/	<sense> ... </sense>

5.9 Implementing the Definition class

The Definition class, which carries a description of a sense, is implemented by means of the TEI element <def>²⁷⁾ (see Table 10).

Table 10 — Definition class

LMF class	TEI construct
/Definition/	<def> ... </def>

27) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-def.html>

5.10 Implementing the OrthographicRepresentation class

The serialization of the OrthographicRepresentation class can be triggered by three classes of the same model: Form, Lemma and Definition. The TEI serialization shall be differentiated by the nature of its subclasses FormRepresentation (see 6.1) and TextRepresentation (see 6.7).

5.11 Implementing the CrossREF class

Two cases shall be differentiated for the implementation of cross references using TEI mechanisms:

- implicit mappings: enabled for all TEI elements by means of the @corresp attribute to point to the intended associated construct;
- explicit mappings: the object being pointed at is indicated by means of a <ref>²⁸⁾ element embedded within an <xr> element inside the appropriate source component in order to point to an internal or external dictionary object (entry, form or sense). In legacy print dictionaries, such a mechanism can be associated with a textual label such as *See*, *See also*, *cf.* The type of the intended relation is indicated by means of the @type attribute on the <xr> element. The lexical type of the target and its notation can be specified by means of, respectively, the @type and @notation attributes on the <ref> element (see Table 11).

Table 11 — CrossREF class

LMF class	TEI construct
/CrossREF/ (implicit)	<code><entry corresp="#entryX"></code> ... <code></entry></code> <code><sense corresp="#senseY"></code> ... <code></sense></code>
/CrossREF/ (explicit)	<code><xr type="synonym"></code> <code><lbl> ... </lbl></code> <code><ref type="entry" target=""> ... <ref></code> <code></xr></code>

Implicit referencing

```

<entry xml:lang="en">
  <form type="lemma">
    <orth>lawful</orth>
  </form>
  <sense xml:id="legal" corresp="#legal-sense1">
    <def>being legal</def>
  </sense>
</entry>

```

28) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-ref.html>

Explicit referencing

```
<entry xml:lang="en">
  <form type="lemma">
    <orth>lawful</orth>
  </form>
  <xr type="synonym">
    <lbl>SYN. see</lbl>
    <ref type="entry" target="#legal">legal</ref>
  </xr>
</entry>
```

The implicit referencing mechanism can be used in the case of multi-word expressions when a form is construed to be made of several sub-forms which can be mapped onto other existing lexical entries. The segmentation of the <form> shall be made by means of the <seg> element, with a @corresp attribute containing a pointer to another entry.

5.12 Data category selection

To establish the link to selected data categories, the TEI Guidelines provide two mechanisms.

For elements having a very precise scope, such as <gen>, <number>, <mood>, two attributes can be activated: @dcr:datcat for data category name, and @dcr:valueDatcat for the corresponding value.

For larger scope TEI elements, such as <form>, @type can replace @dcr:datcat (to carry data category names) for grammatical/syntactic information described within entries.

The following example illustrates the decoration of the representation for gender and number with corresponding data categories.

The namespace associated to the prefix “dcr” shall be conformant to the TEI guidelines.

EXAMPLE

[Inside entries]

```
<gramGrp>
  <gen dcr:datcat="#gender" dcr:valueDatcat="#masculine">masculin</gen>
  <number dcr:datcat="#number" dcr:valueDatcat="#plural">pluriel</number>
</gramGrp>
```

6 Serialization of the MRD model (ISO 24613-2)

6.1 Implementing the form representations for the Form class

When implemented as a subclass of the OrthographicRepresentation class, the FormRepresentation class shall be represented by means of one of the following elements: <orth>, <pron>, <hyph>, <stress>, <syll>, coupled with the appropriate attributes to qualify the content, in particular @type and @notation.

The main two elements that shall be used for indicating the orthographic and the phonetic representation of a form are the following:

- <orth>²⁹⁾ in cases when the transcription is truly orthographic, as defined by the writing convention of the corresponding object language;

29) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-orth.html>

- <pron>³⁰⁾ in cases when the transcription represents the actual pronunciation of the form in any possible notation format.

The two elements can be further refined by means of the @notation attribute to indicate the orthographic (e.g. transliteration scheme) or phonetic (e.g. IPA or SAMPA) notational scheme.

The other possible TEI elements for implementing the orthographic representation of a form are <hyph>³¹⁾ for the hyphenation, <stress>³²⁾ for stress patterns and <syll>³³⁾ for syllabification marks (see Table 12).

Table 12 — FormRepresentation class

LMF class	TEI constructs
/FormRepresentation/	<orth> ... </orth> <pron> ... </pron> <hyph> ... </hyph> <stress> ... </stress> <syll> ... </syll>

EXAMPLE 1 in English [Source: TEI Guidelines]:

```
<entry xml:lang="en">
  <form type="lemma">
    <orth>transcription</orth>
    <pron notation="ipa">trænskriːpʃən</pron>
    <gramGrp>
      <pos>n</pos>
    </gramGrp>
  </form>
</entry>
```

EXAMPLE 2 in Chinese [Source: TEI Guidelines]:

```
<entry xml:lang="zh">
  <form type="lemma">
    <orth>將</orth>
    <pron>jiang4</pron>
  </form>
</entry>
```

EXAMPLE 3 in Arabic

```
<form type="lemma">
  <gramGrp>
    <pos>verb</pos>
  </gramGrp>
  <orth notation="transliterated">kataba</orth>
  <orth notation="arabicUnvowelized">كتب</orth>
  <orth notation="arabicVowelized">كُتِبَ</orth>
</form>
```

30) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pron.html>

31) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-hyph.html>

32) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-stress.html>

33) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-syll.html>

6.2 Implementing classes derived from the Form class

6.2.1 General principles

The TEI Guidelines provide one single element (<form>) for the implementation of the various types of forms identified in ISO 24613-1 and ISO 24613-2. The differentiation between them shall be implemented by means of specific values of the @type attribute.

6.2.2 Implementing the WordForm class

The WordForm class is also derived from the Form class and shall also be implemented by means of the TEI element <form> further constrained using the @type attribute with the value “inflected” (see [Table 13](#)).

Table 13 — WordForm class

LMF class	TEI construct
/FormRepresentation/	<form type="inflected"> ... </form>

6.2.3 Implementing the Stem class

The Stem class is also derived from the Form class for the representation of a stem or a root and shall also be implemented by means of the TEI element <form> further constrained by means of the @type attribute set to “stem” or “root” depending on the linguistic context (see [Table 14](#)).

Table 14 — Stem class

LMF class	TEI constructs
/Stem/	<form type="stem"> ... </form>
	<form type="root"> ... </form>

6.2.4 Implementing the WordPart class

The WordPart class is also derived from the Form class for the representation of sub-lexeme parts and shall also be implemented by means of the TEI element <seg>, which may be further constrained by means of the @type attribute set to “affix”, “infix”, “prefix” or “suffix” depending on the linguistic context (see [Table 15](#)).

Table 15 — WordPart class

LMF class	TEI constructs
/WordPart/	<seg type="affix"> ... </form>
	<seg type="prefix"> ... </form>
	<seg type="suffix"> ... </form>

EXAMPLE

```
<entry xml:lang="en">
  <form type="lemma">
    <orth>
      <seg corresp="#motion">motion</seg>
      <seg corresp="#picture">picture</seg>
    </orth>
    <gramGrp>
      <pos>noun</pos>
    </gramGrp>
  </form>
  <sense>
    <def>sequence of pictures that give the effect of motion when
```

```

    shown in a rapid succession</def>
  </sense>
</entry>

```

6.2.5 Implementing the RelatedForm class

The RelatedForm class is also derived from the Form class to establish the link to related forms. The mapping, which is optional, is activated by means of the TEI @corresp attribute pointing to a target lexical entry representing a derived form of the current one (see [Table 16](#)).

Table 16 — RelatedForm class

LMF class	TEI construct
/RelatedForm/	<form type="relatedForm" corresp="#id"> ... </form>

6.3 Implementing the Bibliography class

The TEI element <bibl>³⁴⁾ shall be used to serialize bibliographic information within the Bibliography class that could be related to another class from the core, the MRD, or the etymological models (see [Table 17](#)).

Table 17 — Bibliography class

LMF class	TEI construct
/Bibliography/	<bibl> ... </bibl>

6.4 Implementing the CrossREF mechanism to refer to external media files

The CrossREF mechanism can be used to make reference to media related information (e.g. audio or video illustrations of an entry or a sense). It shall be implemented in the TEI by means of the <media>³⁵⁾ element.

The <media> element shall be associated with two mandatory attributes (see [Table 18](#)):

- @mimeType to indicate the nature of the designated media. The value of this attribute shall be a valid multimedia internet mail extension (MIME) media type and taken from the corresponding list maintained by IANA³⁶⁾;
- @url to point to the corresponding media file by means of a Uniform Resource Identifier (URI).

Table 18 — Reference to media related information

LMF class	TEI construct
CrossREF for media files	<media mimeType="" url=""> ... </media>

6.5 Implementing the Example class

The Example class provides usage examples of the form within the Sense class. It shall be implemented in the TEI as a construct based on the <cit>³⁷⁾ element. The <cit> element shall contain at least a child <quote>³⁸⁾ (that implements the TextRepresentation class), together with some possible additional

34) <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-bibl.html>

35) <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-media.html>

36) <https://www.iana.org/assignments/media-types/media-types.xhtml>

37) <http://tei-c.org/release/doc/tei-p5-doc/en/html/ref-cit.html>

38) <http://tei-c.org/release/doc/tei-p5-doc/en/html/ref-quote.html>

constraints (see [Table 19](#)). An Example can be further associated to a Translation implemented as described in [6.6](#).

Table 19 — Example class

LMF class	TEI construct
/Example/	<pre><cit type="example"> <quote> ... </quote> ... </cit></pre>

6.6 Implementing the Translation class

The Translation class represents the provision of a translation equivalent within the Sense class. It shall be implemented in the TEI by means of the <cit> element associated to a @type attribute with value "translation" (see [Table 20](#)). As explained in [6.5](#), the relationship with the Example class is reciprocal, which means a serialization of both classes could embed a <cit type="example"> construct within <cit type="translation"> and vice versa.

Table 20 — Translation class

LMF class	TEI construct
/Translation/	<cit type="translation"> ... </cit>

6.7 Implementing the TextRepresentation class

The TextRepresentation class shall be uniformly implemented in TEI using the <quote> element as a child of the <cit> construct to contain the text literal of a Translation or an Example of a Sense (see [Table 21](#)).

Table 21 — TextRepresentation class

LMF class	TEI construct
/TextRepresentation/	<quote> ... </quote>

EXAMPLE

```
<entry xml:lang="fr">
  <form type="lemma">
    <orth>chose</orth>
  </form>
  <sense>
    <cit type="example">
      <quote>Ce qui était une très bonne chose, et fournissait des vêtements à très bas
coût</quote>
    </cit>
    <cit type="translation" xml:lang="de">
      <quote>Sache</quote>
      <gramGrp>
        <gen>f</gen>
      </gramGrp>
      <cit type="example">
        <quote>Was ja auch eigentlich eine gute Sache war, billige Kleidung bereit zu
stellen</quote>
      </cit>
    </cit>
  </sense>
</entry>
```

Additional constraints can cover a whole range of linguistic features, e.g.:

- bibliographical description of the source for the example by means of the <bibl> element;
- grammatical constraints illustrated by the example: <gramGrp>;
- usage information: <usg>;
- translation of the example, by means of a recursive <cit type="translation"> construct.

6.8 Implementing the SubjectField class

The SubjectField class provides domain information for the LexicalEntry, or a specific Sense thereof. It shall be serialized by means of the <usg>³⁹⁾ element together with a @type attribute set to "domain" (see Table 22).

NOTE As stated in 7.2.2, other values for usg/@type are possible when other types of usage information are to be provided.

Table 22 — SubjectField class

LMF class	TEI construct
/SubjectField/	<usg type="domain"> ... </usg>

6.9 Implementing the CrossREF mechanism to represent related entries

As exemplified in ISO 24613-2:2020, B.10.2, the CrossREF mechanism can be used to make reference to entries related to the current one and usually set up in legacy dictionaries as subentries, such as multi-word expressions or idioms, often with short glosses. This can be implemented by means of the <xr> element as described in 5.11, but when it is needed to keep the subordination of the related entry to the main entry, it shall be implemented by means of the <re>⁴⁰⁾ element (see Table 23).

The content model of the <re> element shall comply with the general structure of a normal <entry> as needed to represent the corresponding content.

NOTE This element is equivalent to the use of the recursive <entry> element with the appropriate @type attribute: <entry type="relatedEntry">, as used in the TEI Lex 0 initiative⁴¹⁾.

Table 23 — Related dictionary entries

LMF class	TEI construct
CrossREF for related entries	<re> ... </re>

EXAMPLE (derived from ISO 24613-2, example B.15)

```
<entry>
  <form>
    <orth>picture</orth>
    <gramGrp>
      <pos>noun</pos>
    </gramGrp>
  </form>
  <sense>
    <def>A visual representation made by various means such as painting or
    photography.</def>
  </sense>
</re>
```

39) <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-usg.html>

40) <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-re.html>

41) <https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html>

```

<form>
  <orth>motion picture</orth>
  <gramGrp>
    <pos>noun</pos>
  </gramGrp>
</form>
<sense>
  <def>A sequence of pictures that give the effect of motion when shown in a rapid
  succession.</def>
</sense>
</re>
</entry>

```

7 Implementing the classes from the etymological extension (ISO 24613-3)

7.1 Implementing the Etymology class

The Etymology class encompasses a set of etymological information, for the LexicalEntry in general, or for a specific Sense therein. It shall be implemented in TEI as the <etym> element⁴²⁾. If the user wishes to specify an etymological process on the Etymology, it can be expressed with the @type and @subtype attributes (see Table 24). The possible values for @type on <etym> should be taken from ISO 24613-3:2021, Annex B. Typical values are “borrowing”, “inheritance”, “metaphor”, “metonymy”, “compounding” or “grammaticalization”.

Table 24 — Etymology class

LMF class	TEI construct
/Etymology/	<etym type="..." subtype="..."> ... </etym>

EXAMPLE

```

<etym><date>XIIIe</date> ; languste, <date> v. 1120 </date>, «sauterelle»; <lbl>encore
dans</lbl> Corneille (Hymnes, 7) ; <lang>anc. provençal</lang> langosta, <lbl> altér. du</
lbl> <lang>lat. class.</lang> locusta «sauterelle».</etym>

```

7.2 Implementing the Etymon class

The Etymon class shall be implemented by means of the <cit> element with its @type attribute set to “etymon” (see Table 25).

Table 25 — Etymon class

LMF class	TEI construct
/Etymon/	<cit type="etymon"> ... </cit>

The <cit type="etymon"> element can contain further elements which constitute the description of the current etymon, in particular:

- <form> to provide form related information, as described in 7.2.1;
- <usg> to declare a number of different features about the etymon’s semantic or socio-linguistic usage, as described in 7.2.2;
- <lang>, to encode the explicit descriptive reference to the language associated with an etymon, as described in 7.2.3;
- Further elements such as <gramGrp>, <date>, <bibl> or <ref> to document the etymon as described in 7.2.4 to 7.2.7.

42) <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-etym.html> cf. also Reference [7]

7.2.1 Referencing forms in an etymon

Within an etymon, form content can be embedded within a <form> element with either <orth> or <pron> elements, or both, embedded within the <cit type="etymon">, as in the following example:

EXAMPLE

```
<cit type="etymon">
  <form>
    <pron>...</pron>
    <orth>...</orth>
  </form>
</cit>
```

This element can be associated with an @xml:lang attribute which provides the actual language of the corresponding etymon, and which shall be encoded in accordance with IETF BCP 47.

7.2.2 Representing the meaning of an etymon

The meaning of an etymon can be expressed using the <gloss> (see 7.6) or <def> elements depending on the encoding constraints. <gloss> can be used for a basic summary of the meaning of the etymon, or in the case of bi-/multilingual content for translations. <def> should be used for more detailed descriptions or a definition of the etymon. In both cases, if the <gloss> or <def> are in a language other than that of the etymon, the @xml:lang shall be applied to the given element.

The <usg> element with the @type attribute can be used to declare a number of different features about the etymon's semantic or socio-linguistic usage. It is recommended that the values for @type be compliant with those of the TEI Lex 0 initiative⁴³⁾.

7.2.3 Representing the language of an etymon

The <lang>⁴⁴⁾ element shall be used to encode the explicit descriptive reference to the language associated with an etymon. A @norm attribute can be used to indicate a standardized representation of the corresponding language or language family in compliance with IETF BCP 47.

EXAMPLE <lang expand="Mittelhochdeutsch" norm="gmh">mhd.</lang>

7.2.4 Associating grammatical information to an etymon

The provision of supplemental grammatical information to an etymon shall be made by means of the <gramGrp> element in the same way as for the <form> element (see 5.7).

7.2.5 Dating an etymon

The <date>⁴⁵⁾ element shall be used to mark up the period associated with an etymon, together with the following constraints:

- @type;
- temporal attributes (see below).

Optionally, if the precise date is not known, or the dates concern a span of time, the <date> element can be expressed without text (e.g. <date/>) and the dating information can be specified as attributes. This can be expressed by means of one or more of the following attribute pairs:

- @notBefore, @notAfter;

43) <https://github.com/DARIAH-ERIC/lexicalresources/blob/master/docs/pages/TEILex0/TEILex0.html>

44) <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-lang.html>

45) <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-date.html>