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

**Part 5:
Lexical base exchange (LBX)
serialization**

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

Partie 5: Sérialisation de l'échange de bases lexicales (LBX)



STANDARDSISO.COM : Click to view the full PDF of ISO 24613-5:2022



COPYRIGHT PROTECTED DOCUMENT

© ISO 2022

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 requirements.....	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 OrthographicRepresentation class.....	5
5.7 Implementing the Form class.....	6
5.7.1 Form class.....	6
5.7.2 Lemma class.....	6
5.8 Implementing the GrammaticalInformation class.....	6
5.9 Implementing the Sense class.....	7
5.10 Implementing the Definition class.....	8
5.11 Implementing the CrossREF class.....	8
6 Serialization of the MRD extension (ISO 24613-2).....	10
6.1 Implementing OrthographicRepresentation subclasses.....	10
6.2 Implementing the FormRepresentation class.....	10
6.3 Implementing the Form subclasses.....	11
6.3.1 General principles.....	11
6.3.2 Implementing the WordForm class.....	11
6.3.3 Implementing the Stem class.....	11
6.3.4 Implementing the WordPart class.....	11
6.3.5 Implementing the RelatedForm class.....	12
6.3.6 Implementing the TextRepresentation class.....	13
6.3.7 Implementing the Translation class.....	14
6.3.8 Implementing the Example class.....	14
6.4 Implementing the SubjectField class.....	15
6.5 Implementing the Bibliography class.....	15
7 Implementing the CrossREF mechanism to refer to external media files.....	15
8 Implementing the classes from the etymological extension (ISO 24613-3).....	15
8.1 Implementing the Etymology class.....	15
8.2 Implementing the Etymon class.....	16
8.2.1 General.....	16
8.2.2 Referencing forms in an etymon.....	16
8.2.3 Representing the meaning of an etymon.....	16
8.2.4 Representing the language of an etymon.....	16
8.2.5 Dating an etymon.....	17
8.2.6 Providing sources associated with an etymon.....	17
8.3 Implementing the EtyLink class.....	17
8.4 Implementing the CognateSet class.....	17
8.5 Implementing the Cognate class.....	17
9 Additional mechanisms.....	18
9.1 Overview.....	18
9.2 XML feature structure implementation.....	18
9.3 Representing various labels with <LBL>.....	18
9.4 Providing rendering information with the @rend attribute.....	18

Annex A (informative) LBX data category selection	19
Annex B (informative) LBX feature structure implementation	24
Annex C (informative) LBX examples for applying LBX serialization	27
Bibliography	32

STANDARDSISO.COM : Click to view the full PDF of ISO 24613-5:2022

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 24613-5, together with ISO 24613-1:2019, ISO 24613-2:2020, ISO 24613-3:2021 and ISO 24613-4:2021, cancels and replaces ISO 24613:2008, which has been technically revised.

The main change compared to the previous edition is as follows:

- entire revision of the content and its subdivisions 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.

STANDARDSISO.COM : Click to view the full PDF of ISO 24613-5:2022

Language resource management — Lexical markup framework (LMF) —

Part 5: Lexical base exchange (LBX) serialization

1 Scope

This document describes the serialization of the lexical markup framework (LMF) model defined as an extensible markup language (XML) model derived from the language base exchange (LBX) schema and compliant with the W3C XML schema. This serialization covers the classes, data categories, and mechanisms of ISO 24613-1 (core model), ISO 24613-2 (machine-readable dictionary (MRD) model), and ISO 24613-3 (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 15924, *Information and documentation — Codes for the representation of names of scripts*

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. PHILLIPS, A., DAVIS, M. (eds.), September 2009. Best Current Practice. Available from: <https://tools.ietf.org/html/bcp47>

W3C. Extensible Markup Language (XML) 1.1 (Second Edition). W3C Recommendation 16 August 2006, edited in place 29 September 2006. Available from: <https://www.w3.org/TR/2006/REC-xml11-20060816/>

3 Terms and definitions

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

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

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

4 General requirements

This document aims at providing constructs for each LMF class from ISO 24613-1 (core model), ISO 24613-2 (MRD extension), and ISO 24613-3 (etymological extension). It requires compliance

with ISO 24613-1, ISO 24613-2, and ISO 24613-3 when implementing data categories referred to in the respective parts, and compliance with the W3C XML Schema 1.1 for representing structured information in XML. LBX extends the original models by means of data category selections and precise value lists, the creation of new subclasses and the definition of new constraints. In addition, this document complies with the cardinalities expressed in ISO 24613-1, ISO 24613-2, and ISO 24613-3. The LBX serialization is richer in detail than LMF, in order to meet specific design objectives. Still, this document does not elaborate on the metadata aspects from LMF, since the LBX schema is by essence much richer for the representation of all the aspects related to the creation, content, versioning and database implementation of lexical content at large. Occasionally, slightly equivalent constructs to explicit requirements from the LMF standard are mentioned.

The XML examples in this document are simplified by omitting namespaces. Except where otherwise stated, it is assumed that XML elements belong to the LBX namespace and that the examples lie within the scope of the following XML namespace declaration:

```
xmlns="http://www.LexicalBaseExchange.org/2021/schema"
```

Besides, datatypes in this document are defined in compliance to the XML Schema Part 2 recommendation. The "xs:" prefix corresponds to the following namespace:

```
http://www.w3.org/2001/XMLSchema
```

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

5.1 Implementing the LexicalResource class

The LexicalResource class shall be implemented in LBX by means of the <LexicalResource> element (see [Table 1](#)), which groups together one to many lexicons in a single collection. This level may be omitted in cases where the lexical resource contains only one lexicon so that the resource starts directly with the lexicon level. In cases where a lexical resource contains a large number of lexicons or several very large lexicons, the lexicon (XML document) can reference a virtual lexical resource using a @lexicalResourceID in the <Lexicon> element and optionally the <LexicalEntry> element (see [5.5](#)).

Table 1 — LexicalResource class

LMF class	LBX construct
/LexicalResource/	<LexicalResource>

5.2 Implementing the GlobalInformation class

The GlobalInformation class shall be implemented in LBX by means of the <GlobalInformation> element (see [Table 2](#)) either by referencing a GlobalInformation.xsd schema using an <xsd:include> element, or as a direct child of a <LexicalResource> element. <GlobalInformation> allows the encoding of a variety of administrative, technical, documentary, and bibliographic information attached to the corresponding lexical resource.

Table 2 — GlobalInformation class

LMF class	LBX construct
/GlobalInformation/	<GlobalInformation>

Since the LBX serialization is based on the W3C recommendation for XML, it implements the @xml:lang attribute to indicate language information corresponding to the content of specific elements. According to the W3C recommendation, @xml:lang content shall be compliant with BCP 47. There is no need for a specific implementation of the /language coding/ data category or the /script coding/ data category in order to ensure compliance of this document with ISO 24613-1. LBX does allow the inclusion of these data categories in the <GlobalInformation> element in order to support the validation of equivalent metadata found in the <LexiconInformation> elements

of one or more lexicons (see 5.4). When included, the /script coding/ shall use the codes from ISO 15924. The /character encoding/ data category is implemented in the XML declaration of an LBX conformant document using the @encoding attribute. For instance, an XML-LBX document encoded as UTF-8 according to the Unicode standard shall begin with the following declaration:

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

A non-exclusive list of <GlobalInformation> sub-elements, simple types indexed by value, follows:

- “ISO 639-3”, a simple type enumerating the set of language codes used across all lexicons;
- “ISO 15924”, a simple type enumerating the set of scripts used across all lexicons;
- GlobalNotationType, a simple type enumerating the set of notations used across all lexicons;
- GlobalPartOfSpeechType, a simple type enumerating the set of <partOfSpeech> values used across all lexicons;
- SubjectFieldType, a simple type enumerating the set of <SubjectField> values used across all lexicons.

Examples can be found in the LBX reference schema, GlobalInformation document (see Annex B).

5.3 Implementing the Lexicon class

The Lexicon class shall be implemented in LBX by means of the <Lexicon> element (see Table 3), which is a direct child of the <LexicalResource> element when <LexicalResource> is used. If the <LexicalResource> element is not used, <Lexicon> becomes the root element. In cases where a lexical resource contains a large number of lexicons or several very large lexicons, the lexicon (XML document) can reference a virtual lexical resource using a @lexicalResourceID in the <Lexicon> element (see 5.1). In the case of a virtual lexical resource, where the <LexicalResource> element is not part of the same XML document as the <Lexicon> element, the lexicon can use an include statement to reference a relevant <GlobalInformation> element. Other information within the <Lexicon> element should be qualified through the following child element(s) and attributes as direct children of the <Lexicon> element or, optimally, as children of the <LexiconInformation> element (see 5.4):

- <Title>, the title of the lexicon;
- @lexiconID, of datatype xs:ID as a unique identifier for the lexicon; as a best practice, the id should be a URI and be unique within a language resource; @xml:ID can be used in place of @lexiconID when there is a design intent to make the entry accessible on the web;
- @lexicalResourceID of datatype xs:ID as a unique identifier for the lexical resource; as a best practice, the ID should be a URI for global scope; in addition, @xml:ID can be used in place of @lexicalResourceID when there is a design intent to make the entry accessible on the web;
- @lexiconType, of datatype "xs:string"; the type of lexicon, e.g. bilingual dictionary, monolingual dictionary;
- @sourceLanguage, of datatype "xs:string"; the language of the <Lemma> element or its inflected forms;
- @targetLanguage, of datatype "xs:string"; the language the lemma is translated to, principally represented in the <Translation> element.

Table 3 — Lexicon class

LMF class	LBX construct
/Lexicon/	<Lexicon>

5.4 Implementing the LexiconInformation class

The LexiconInformation class shall be implemented in LBX by means of the <LexiconInformation> element (see [Table 4](#)) either by referencing a LexiconInformation.xsd schema using an <xsd:include> element or as a direct child of the <Entry> element. <LexiconInformation> allows the encoding of a variety of administrative, technical, documentary, and bibliographic information attached to the corresponding lexical entry.

Table 4 — LexiconInformation class

LMF class	LBX construct
/LexiconInformation/	<LexiconInformation>

When not included in the <Lexicon> element, information qualifying the lexicon shall be included as elements and attributes in the <LexiconInformation> element. These include (see [5.3](#)):

- <Title>;
- @lexiconID;
- @lexicalResourceID;
- @lexiconType;
- @sourceLanguage;
- @targetLanguage.

The <LexiconInformation> can also include elements and data categories that further qualify information in the lexicon and can be used to support the validation of the XML document (lexicon). These elements and data categories should also be included in the global set of elements and data categories found in the <GlobalInformation> element (see [5.2](#)) and a comparison of the corresponding values in <GlobalInformation> and <LexiconInformation> should be part of the validation process.

A non-exclusive list of these sub-elements, simple types indexed by value, follows:

- NotationType, a simple type enumerating the set of notations used in a lexicon;
- PartOfSpeechType, a simple type enumerating the set of <partOfSpeech> values used in a lexicon;
- SubjectFieldType, a simple type enumerating the set of <SubjectField> values used in a lexicon.

NOTE In addition to the <LexiconInformation> construct, LBX allows the concatenation of lexicon information for a subset of lexicons grouped by language by referencing a named language data schema (e.g. ArabicLanguageData.xsd) (see [Clause B.1](#)).

5.5 Implementing the LexicalEntry class

The LexicalEntry class shall be implemented in LBX by means of the <Entry> element (see [Table 5](#)). Lexical information inside <Entry> elements should be encoded through the following child elements:

- <GramFeats> for grammatical information related to the whole entry;
- <Form> for containing the text literal and attributes qualifying the text literal (the Form class is serialized through subclasses in LBX);
- <Etymology> for etymological aspects;
- <Sense> for semantic information;
- <Xref> for referencing internal or external elements.

Attributes used for the <LexicalEntry> element can include:

- @entryID of datatype xs:ID as a unique identifier for an entry; as a best practice, the id should be a URI and be unique within a language resource; @xml:ID can be used in place of @entryID when there is a design intent to make the entry accessible on the web;
- @lexiconID of datatype xs:ID as a unique identifier for the parent lexicon; as a best practice, the id should be a URI and be unique within a language resource; @xml:ID can be used in place of @entryID when there is a design intent to make the lexicon accessible on the web;
- @lexicalResourceID, a reference to the @lexicalResourceID of the associated lexicon collection when there is more than one lexicon.

Table 5 — LexicalEntry class

LMF class	LBX construct
/LexicalEntry/	<Entry>

The following example in French illustrates the encoding of a simple dictionary entry with two senses.

EXAMPLE

```
<Entry xml:lang="fr">
  <Etymology>XIIIe; languste, v. 1120, «sauterelle»; encore dans Corneille (Hymnes, 7);
  anc. provençal langosta, altér. du lat. class. locusta «sauterelle».</Etymology>
  <Lemma>
    <GramFeats>
      <POS>noun</POS>
      <Gender>fem</Gender>
    </GramFeats>
    <FormRep xml:lang="fr" notation="French">langouste</FormRep>
    <FormRep xml:lang="fr" notation="IPA">lägust</FormRep>
  </Lemma>

  <Sense senseNR="1">
    <Def>
      <DefRep xml:lang="fr">Grand crustacé marin (Décapodes macroures) aux pattes
      antérieures dépourvues de pinces, aux antennes longues et fortes, et dont la chair est
      très appréciée.</DefRep>
    </Def>
  </Sense>

  <Sense senseNR="2">
    <Note type="socioCultural">Fig. et fam. (vulg.).</Note>
    <Def>
      <DefRep xml:lang="fr">Femme, maîtresse</DefRep>
    </Def>
  </Sense>
</Entry>
```

NOTE 1 The style in the above example is appropriate for use in a lexical resource that contains a collection of bilingual lexicons in a variety of source languages, e.g. French, Spanish, Russian, Chinese. A simpler style can be used for a collection of monolingual French lexicons. For example, <Orth> and <Pron> can be used in place of the equivalent <FormRep> elements and the <Def> element can directly contain the text content rather than employing a <DefRep> child element for managing text content (see 5.10). See 6.2 for an example of simplification using the <Orth> and <Pron> elements.

NOTE 2 The @notation value “French” is short for “Canonical French”.

5.6 Implementing the OrthographicRepresentation class

Classes containing an OrthographicRepresentation class include the Form, Lemma, and Definition classes. Orthographic representations shall be implemented in LBX by means of elements corresponding to OrthographicRepresentation subclasses that are introduced in ISO 24613-2 (machine-readable dictionary (MRD) model), or possible new OrthographicRepresentation subclasses derived through the

principles for LMF extensions described in ISO 24613-1 (core model). ISO 24613-1:2019, 5.6.1, describes some of the representation types that can serve as a basis for extending the OrthographicRepresentation class. ISO 24613-4:2021 (TEI extension), 6.1, lists a number of representation elements that are valid for use with the Form class. Elements implemented in this part are described in 5.7.2, 5.10, and successive subclauses from 6.3.2 to 6.3.8.

5.7 Implementing the Form class

5.7.1 Form class

The Form class shall be implemented in LBX by elements that instantiate Form subclasses (see Table 6, 6.2 and 6.3).

Table 6 — Form class

LMF class	LBX construct
/Form/	<Form>

5.7.2 Lemma class

The Lemma class, a subclass of the Form class, shall be implemented in LBX by means of the <Lemma> element (see Table 7).

Table 7 — Lemma class

LMF class	LBX construct
/Lemma/	<Lemma>

Orthographic representations in the <Lemma> element shall be implemented in LBX by means of the <FormRep> element, or by elements that instantiate Form subclasses, including <Orth> and <Pron>.

NOTE 1 The <FormRep>, <Orth>, and <Pron> elements are introduced in 6.2.

NOTE 2 <Orth> and <Pron> can be allowed when justified by design goals.

5.8 Implementing the GrammaticalInformation class

The GrammaticalInformation class groups grammatical features associated with the LexicalEntry class, Form class, or other classes (e.g. Translation, Sense) in case of specific grammatical restrictions. The GrammaticalInformation class shall be implemented in LBX by means of the <GramFeats> element (see Table 8) combined with various possible child elements for specific grammatical features.

Table 8 — GrammaticalInformation class

LMF class	LBX construct
/GrammaticalInformation/	<GramFeats>

LBX provides the following child elements of <GramFeats> for describing specific grammatical features of associated elements (e.g. <Lemma>, <WordForm>):

- <POS> to indicate the grammatical category of the lexical item. This corresponds to the /partOfSpeech/ data category in ISO 24611:2012, Annex A;
- <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;
- <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;

- <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;
- <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;
- <Aspect> to indicate the grammatical aspect (if relevant) of the lexical item or one of its inflected forms;
- <Mood> to indicate the grammatical mood (if relevant) of the lexical item or one of its inflected forms;
- <Voice> to indicate the grammatical voice (if relevant) of the lexical item or one of its inflected forms;
- <Animacy> to indicate the grammatical animacy (if relevant) of the lexical item or one of its inflected forms (e.g. in Russian);
- <GrammaticalClass> to indicate the grammatical class (gender) of Bantu languages;
- <GrammaticalClassGroup> to indicate the aggregate grammatical classes (genders) of a specific noun in the singular and plural;
- <iType> to indicate the inflectional class associated with the lexical item or one of its inflected forms;
- <Subcat> to indicate subcategorization information (e.g. transitive/intransitive, countable/non-countable).

The following example shows the grammatical information for a word form in a monolingual French dictionary that is part of a notional language resource containing a collection of monolingual and bilingual dictionaries in multiple source languages. The @notation="French", denoting canonical French, is used in databases that support a large set of possibly idiosyncratic notations (e.g. for canonical, transliterated and transcribed forms).

EXAMPLE

```
<Entry>
  <Lemma>
    <GramFeats>
      <POS>verb</POS>
      <Subcat>transitive</Subcat>
    </GramFeats>
    <FormRep xml:lang="fr" notation="French">pacifier</FormRep>
    <FormRep xml:lang="fr" notation="ipa">pasifje</FormRep>
  </Lemma>
  <Sense/>
</Entry>
```

For an example of simplifying this schema, see [6.2](#).

5.9 Implementing the Sense class

The Sense class, as a recursive construct, shall be implemented in LBX by means of the <Sense> element (see [Table 9](#)). LBX does not allow character content in the element.

Table 9 — Sense class

LMF class	LBX construct
/Sense/	<Sense>

5.10 Implementing the Definition class

The Definition class, which contains a narrative description of the word sense, shall be implemented in LBX by means of the <Def> element (see [Table 10](#)).

Table 10 — Definition class

LMF class	LBX construct
/Definition/	<Def>

LBX provides the following child element for the description of definition information:

- <DefRep>, an element instantiating a TextRepresentation subclass containing the character content of the definition. See [6.3.6](#).

The <Def> element allows mixed data enabling the text literal (character content) to be contained within the <Def> element itself. See [6.3.6](#) for a description of the <DefRep> element, which provides an alternative approach for managing the text literal. Within an LBX <LexicalResource> or <Lexicon> element, the consistent use of <Def> or <DefRep> for character content is a best practice.

NOTE The <DefRep> element supports the inclusion of multiple orthographic representations for a <Def> element (e.g. Simplified Chinese, Traditional Chinese).

5.11 Implementing the CrossREF class

The CrossREF class shall be implemented in LBX by means of the <Xref> element (see [Table 11](#)), which points to an internal or external dictionary object, such as an entry, lemma, sense, or translation. LBX allows a range of different data types for the cross reference (URI, IRI, HREF, etc.). In order to make the data accessible through the web, LBX should implement web standards, such as a URL or Resource Description Framework (RDF). The <Xref> element can be qualified through attributes, such as @relType for describing the relationship type (e.g. synonym, antonym, hyponym). The type of identifier and any of its inherent characteristics and constraints should be identified in the <GlobalInformation> element or the <LexiconInformation> element, as appropriate.

Table 11 — CrossREF class

LMF class	LBX construct
/CrossREF/	<Xref [URI IRI HREF ...]="">

The <Xref> element can be used as a child of the <RelForm> element to point to content and metadata in a different entry. In this case, the <Xref> element can replace the content and metadata contained in the <RelForm>.

EXAMPLE 1

```
<Entry>
  <Lemma>
    <FormRep xml:lang="en">lawful</FormRep>
  </Lemma>
  <RelForm relType="synonym">
    <Xref href="#legal">legal</Xref>
  </RelForm>
  <Sense/>
</Entry>
```

The <Xref> element can also implement the CrossREF class through other classes. As in the following example, linking related senses provides a better description of semantic relationships.

EXAMPLE 2

```

<Entry>
  <Lemma>
    <FormRep xml:lang="en">lawful</FormRep>
  </Lemma>
  <Sense xml:id="legal">
    <Def>being legal</Def>
    <Xref IRI="#legal-sense1"/>
  </Sense>
</Entry>

```

LBX allows multiple strategies for describing multi-word expressions using the CrossREF mechanism. In the following example, the <Xref> element is contained in the Lemma and used to point to other entries, each of which contains a component of the multi-word expression.

EXAMPLE 3

```

<Entry>
  <Lemma formStructure="MWE">
    <GramFeats>
      <POS>noun</POS>
    </GramFeats>
    <FormRep xml:lang="en">motion picture</FormRep>
    <Xref relType="component" order="1" IRI="#motion_form_1">motion</Xref>
    <Xref relType="component" order="2" IRI="#picture_form_1">picture</Xref>
  </Lemma>
  <Sense>
    <Def>sequence of pictures that give the effect of motion when shown in rapid
    succession</Def>
  </Sense>
</Entry>

```

In the following example, the use of the <RelForm> to implement the multi-word expression allows a more in-depth grammatical analysis.

EXAMPLE 4

```

<Entry>
  <Lemma formStructure="MWE">
    <GramFeats>
      <POS>noun</POS>
    </GramFeats>
    <FormRep xml:lang="en">motion picture</FormRep>
  </Lemma>
  <RelForm relType="MWEComponent">
    <GramFeats>
      <POS>attributiveNoun</POS>
    </GramFeats>
    <FormRep xml:lang="en">motion</FormRep>
    <Xref relType="component" order="1" IRI="#motion_form_1"/>
  </RelForm>
  <RelForm relType="MWEComponent">
    <FormRep xml:lang="en">picture</FormRep>
    <Xref relType="component" order="2" IRI="#picture_form_1"/>
  </RelForm>
  <Sense>
    <Def xml:lang="en">sequence of pictures that give the effect of motion when shown
    in rapid succession</Def>
  </Sense>
</Entry>

```

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

6.1 Implementing OrthographicRepresentation subclasses

The OrthographicRepresentation class in LBX shall be serialized by means of elements derived from the FormRepresentation and TextRepresentation classes or their subclasses described in ISO 24613-2. FormRepresentation subclasses are further described in 6.2 and TextRepresentation subclasses in 6.3.6. In all cases, these corresponding elements can be qualified by attributes, including @xml:lang, @script, and @notation. Other attributes, such as @representationType (e.g. canonicalForm, phoneticForm), are also available. An LBX implementation should use BCP 47 for language description, especially for the support of web-based applications, data interchange, and system interoperability (@script is not used when BCP 47 is implemented).

6.2 Implementing the FormRepresentation class

Depending on design goals, the FormRepresentation class in LBX shall be serialized by means of a general <FormRep> element derived from the FormRepresentation class, or by elements derived from FormRepresentation subclasses that correspond to the lexical environment of the Form subclasses (see Table 12). The goal of this design is to support effective resource management for large-scale, complex lexical databases (e.g. many lexicons encompassing many languages). When justified by design goals, simplification can be achieved by reducing the number of subclasses employed. These subclasses are represented by the elements described in 6.3.1 to 6.3.5, coupled with the appropriate attributes to qualify the content, in particular @xml:lang, @script, and @notation.

Table 12 — FormRepresentation class

LMF class	LBX construct
/FormRepresentation/	<FormRep> <StemRep> <PartRep> <RelFormRep>

<FormRep> is contained by the <Lemma> (see 5.7.2), <WordForm>, and <RelForm> elements; <StemRep>, <PartRep>, and <RelFormRep> are restricted to specific elements (see 6.3, 6.4 and 6.5).

Where the FormRepresentation class by itself is sufficient without further qualification, equivalents of TEI elements, such as <Orth>, <Pron>, <Hyph>, <Stress>, and <Syll> can potentially be used for further simplification. In such cases, descriptions of qualifying attributes (e.g. @xml:lang) should be included in <GlobalInformation> or <LexiconInformation>, as appropriate. The following example shows the <Orth> and <Pron> elements for a lemma in a monolingual French dictionary, part of a collection of monolingual French lexicons. In reference to ISO 24613-4:2021, 5.2, there is no requirement to include the language and script codes in the <GlobalInformation> element (although LBX allows the inclusion of these codes). <GlobalInformation> can also be used to implement a definition that assigns the "ipa" @notation value to the <Pron> element. The following example shows the implementation of this principle using a revised version of the example in 5.8.

EXAMPLE

```
<Entry>
  <Lemma>
    <GramFeats>
      <POS>verb</POS>
      <Subcat>transitive</Subcat>
    </GramFeats>
    <Orth>pacifier</Orth>
    <Pron>pasifje</Pron>
  </Lemma>
  <Sense/>
</Entry>
```

6.3 Implementing the Form subclasses

6.3.1 General principles

Form subclasses described in ISO 24613-2 shall be serialized by means of the elements described in 6.3.2 to 6.3.5. LBX typically treats the Form class described in ISO 24613-1 as an abstract class.

6.3.2 Implementing the WordForm class

The WordForm class, a subclass of the Form class, shall be implemented in LBX by means of the <WordForm> element (see Table 13). The <WordForm> element can further be characterized by a @formType attribute (e.g. inflection, abbreviation, etc.) and other qualifying values (see Annex B).

Table 13 — WordForm class

LMF class	LBX construct
/WordForm/	<WordForm>

Orthographic representation in the <WordForm> element should be encoded through the following child elements:

- <FormRep>;
- <Orth>;
- <Pron>.

NOTE The FormRep derived elements <Orth> and <Pron> can be used in limited contexts when warranted by design goals (see 6.2).

6.3.3 Implementing the Stem class

The Stem class is derived from the Form class for the representation of a stem or root, and shall be implemented by the <Stem> element (see Table 14) further constrained by means of the @stemType attribute (stem, root, arabicRoot, etc.) depending on the linguistic context.

Table 14 — Stem class

LMF class	LBX construct
/Stem/	<Stem>

Orthographic representation in the <Stem> element should be encoded through the following child element:

- <StemRep>.

6.3.4 Implementing the WordPart class

The WordPart class, a subclass of the Form class, shall be implemented in LBX by means of the <WordPart> element (see Table 15). <WordPart> represents a sub-lexeme component of a word form that is not a stem or root (e.g. prefix, suffix).

Table 15 — WordPart class

LMF class	LBX construct
/WordPart/	<WordPart>

Orthographic representation in the <WordForm> element should be encoded through the following child element:

— <PartRep>.

Examples of use cases include filling prefix and suffix slots in agglutinative languages and “indexing” lexical items by means of shared affixes. The following example illustrates the latter use case.

EXAMPLE

```
<Entry>
  <Lemma>
    <FormRep xml:lang="de">entdecken</FormRep>
  </Lemma>
  <WordPart>
    <PartRep>ent</PartRep>
    <Xref IRI="#ent-prefix"/>
  </WordPart>
  <Sense>
    <Def xml:lang="en">to discover</Def>
  </Sense>
</Entry>
<Entry>
  <Lemma>
    <FormRep xml:lang="de">empfehlen</FormRep>
  </Lemma>
  <WordPart>
    <PartRep>emp</PartRep>
    <Xref IRI="#ent-prefix"/>
  </WordPart>
  <Sense>
    <Def xml:lang="en">to recommend</Def>
  </Sense>
</Entry>
<Entry entryID="ent-prefix">
  <Lemma partType="prefix">
    <FormRep xml:lang="de">ent</FormRep>
  </Lemma>
  <WordForm partType="prefix" formType="variant">
    <FormRep xml:lang="de">emp</FormRep>
  </WordForm>
  <Sense>
    <Def xml:lang="en">An inseparable verbal prefix originally denoting the beginning
of an action or separation. This fundamental meaning has been lost in many verbs.</Def>
  </Sense>
</Entry>
```

6.3.5 Implementing the RelatedForm class

The RelatedForm class is derived from the Form class and shall be implemented in LBX by means of the <RelForm> element (see [Table 16](#)). The <RelForm> element contains a word, phrase, or partial dictionary entry related to the Lemma, but not necessarily to the Sense. The <RelForm> element can contain an optional CrossREF used to establish a link to related forms in other entries (antonym, synonym, component of multi-word expression, etc.). When an equivalent of the related form is not found in another entry, the <RelForm> represents a “degenerate” entry (in which case there is no cross reference). The <RelForm> entry can contain a word form or phrase, a mapping to a different lexical entry (CrossREF), or both a word form and a mapping.

Table 16 — RelatedForm class

LMF class	LBX construct
/RelatedForm/	<RelForm>

Orthographic representation in the <RelForm> element should be encoded through the following child elements:

- <FormRep>; or
- <RelFormRep>.

The following example illustrates the most common use of <RelForm> in LBX. This design approach is commonly used to capture information found in print dictionaries. Two solutions for improving a digital version include to use an <Xref> to map to a new entry for “United Nations”, or better, map back to component entries from a new entry containing a multi-word expression (“United Nations”).

EXAMPLE

```
<Entry>
  <Lemma>
    <FormRep>united</FormRep>
  </Lemma>
  <RelForm>
    <FormRep xml:lang="en">United Nations</FormRep>
    <Gloss xml:lang="en">An international organization for the promotion of
international cooperation and peace</Gloss>
  </RelForm>
  <Sense>
    <Def xml:lang="en">Joined together politically, socially, or economically for a
common purpose.</Def>
  </Sense>
</Entry>
```

6.3.6 Implementing the TextRepresentation class

The TextRepresentation class shall be uniformly implemented in LBX using the <Quote> element (see [Table 17](#)) as a construct to contain the text literal of a Translation or an Example of a Sense. The TextRepresentation class can be implemented in LBX as an optional construct to contain the text literal of a Definition. Alternatively, the <Def> element allows mixed data, enabling the text literal (character content) to be contained within the <Def> element itself (see [5.10](#)). Within an LBX LexicalResource or Lexicon, the consistent use of <Def> or <DefRep> for character content is a best practice.

NOTE The <DefRep> element supports the inclusion of multiple orthographic representations for a <Def> element (e.g. Simplified Chinese, Traditional Chinese).

Table 17 — TextRepresentation class

LMF class	LBX construct
/TextRepresentation/	<TextRep> <DefRep> (with <Def>) <Quote> (with <Translation>, <Example>)

EXAMPLE

```
<Entry>
  <Lemma>
    <GramFeats>
      <POS>noun</POS>
      <Gender>feminine</Gender>
    </GramFeats>
    <FormRep xml:lang="fr">chose</FormRep>
  </Lemma>
  <Sense senseID="chose_sense1" relType="synonym" IRI="#truc_sense1">
    <Example>
      <Quote xml:lang="fr" quoteID="chose_quote1">Ce qui était une très bonne
chose, et fournissait des vêtements à très bas coût</Quote>
    </Example>
  </Sense>
</Entry>
```

```

<Entry>
  <Lemma>
    <GramFeats>
      <POS>noun</POS>
      <Gender>feminine</Gender>
    </GramFeats>
    <FormRep xml:lang="de">Sache</FormRep>
  </Lemma>
  <Sense senseID="truc_sense1" relType="synonym" IRI="#chose_sense1">
    <Example>
      <Quote xml:lang="de" quoteID="truc_quot1">Was ja auch eigentlich eine gute
      Sache war, billige Kleidung bereit zu stellen.</Quote>
    </Example>
  </Sense>
</Entry>

```

Additional constraints can cover a whole range of linguistic features, for example:

- bibliographical description of the source for the example by means of the <Bibl> element;
- grammatical constraints illustrated by the example: <GramFeats>;
- usage information implemented using elements or data categories;
- translation of the example, by means of a <Quote> construct.

6.3.7 Implementing the Translation class

6.3.7.1 General principles

The Translation class represents a translation equivalent of the Lemma. It shall be implemented in LBX by the <Translation> element (see [Table 18](#)), which has a mixed content model accepting both text nodes and sub-elements. The <Translation> element can be qualified by @xml:lang and other relevant attributes.

Table 18 — Translation class

LMF class	LBX construct
/Translation/	<Translation>

Orthographic representation in the <Translation> element should be encoded through the following child element:

- <Quote>.

6.3.7.2 Implementing the gloss construct

LBX qualifies the Translation class using a @type value of “gloss” in order to provide a translation for a <Quote>, <RelForm>, or <RelFormRep> element. LBX typically implements the <Translation type="gloss"> element as an equivalent <Gloss> element, which can be further qualified by @xml:lang and other relevant attributes.

6.3.8 Implementing the Example class

The Example class shall be serialized in LBX by means of the <Example> element (see [Table 19](#)). It contains examples of usage illustrating a particular sense of the Form.

Table 19 — Example class

LMF class	LBX construct
/Example/	<Example>

Orthographic representation in the <Example> element should be encoded through the following child element:

- <Quote>.

A translation of the source language in the <Quote> element can be encoded through the following child element:

- <Gloss>.

6.4 Implementing the SubjectField class

The SubjectField class provides domain information for a specific Sense of a <LexicalEntry>. The SubjectField class shall be serialized in LBX by means of the <Domain> element (see [Table 20](#)).

Table 20 — SubjectField class

LMF class	LBX construct
/SubjectField/	<Domain>

6.5 Implementing the Bibliography class

The LBX element <Bibl> (see [Table 21](#)) shall be used to serialize bibliographic information within the Bibliography class that can be related to another class from the core, the MRD, or the Etymology meta-models.

Table 21 — Bibliography class

LMF class	LBX construct
/Bibliography/	<Bibl>

7 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 LBX by means of the <Media> element (see [Table 22](#)).

The <Media> element shall be associated with two mandatory attributes:

- @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 taken from the corresponding list maintained by the IANA;
- @url to point to the corresponding media file by means of an Internationalized Resource Identifier (IRI) or Uniform Resource Identifier (URI).

Table 22 — Reference to media related information

LMF class	LBX construct
/CrossREF/ for media files	<Media mimeType="" url="">

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

8.1 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 LBX as the <Etymology> element (see [Table 23](#)).

If the users wish to specify an etymological process on the Etymology, it can be expressed with the @etymType attribute. The possible values for @etymType on <Etymology> should be taken from ISO 24613-3:2021, Annex B. Typical values are “borrowing”, “inheritance”, “metaphor”, “compounding”, or “grammaticalization”.

Table 23 — Etymology class

LMF class	LBX construct
/Etymology/	<Etymology etymType="">

8.2 Implementing the Etymon class

8.2.1 General

The Etymon class shall be implemented in LBX by means of the <Etymon> element (see [Table 24](#)). The element can contain a further element which constitutes the description of the current element.

Table 24 — Etymon class

LMF class	LBX construct
/Etymon/	<Etymon>

8.2.2 Referencing forms in an etymon

Within an etymon, form content can be embedded within a <Form> element containing relevant <FormRep> elements, as in the following example:

EXAMPLE

```
<Etymon>
  <Form>
    <FormRep xml:lang="" notation=""></FormRep>
  </Form>
</Etymon>
```

These elements 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 BCP 47. In cases called for by design objectives, the elements <Orth> and <Pron> can be used in place of <FormRep>.

8.2.3 Representing the meaning of an etymon

The meaning of an etymon can be expressed using the <Def> element. If the <Def> element is in a language or languages other than that of the etymon, the @xml:lang attribute shall be applied to the relevant <DefRep> element.

An appropriate @etymonType attribute can be used to declare a number of different features about the element's semantic or socio-linguistic usage.

8.2.4 Representing the language of an etymon

The <Lang> element shall be used to encode the explicit descriptive reference to the language associated with an etymon. A @norm attribute can be used to point to a standard representation of the corresponding language or language family in compliance with BCP 47.

8.2.5 Dating an etymon

The <Date> element shall be used to mark up the period associated with the 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 character content and the information can be specified by attributes. This can be expressed by means of one or more of the following attribute pairs:

- @notBefore; @notAfter;
- @from; @to;
- @when; @precision.

8.2.6 Providing sources associated with an etymon

Bibliographic sources can be cited within an etymology or an etymon using the <Bibl> element. If the bibliography contains pointers to external sources, the CrossREF element <Xref> can be used.

8.3 Implementing the EtyLink class

In case the etymological relation cannot be expressed just by providing a @type attribute on the <Etym> element, this document provides a specific implementation of the EtyLink class from ISO 24613-3 by means of the <Xref> element with a @refType attribute set to "etyLink". The <Xref> element can be further characterized by other attributes describing etymological features such as temporal sequence.

8.4 Implementing the CognateSet class

The CognateSet class shall be implemented in LBX by means of the <CognateSet> element (see [Table 25](#)). The element can contain a further element which constitutes the description of the current etymon.

Table 25 — CognateSet class

LMF class	LBX construct
/CognateSet/	<CognateSet>

8.5 Implementing the Cognate class

The Cognate class shall be implemented in LBX by means of the <Cognate> element (see [Table 26](#)). The <Cognate> element can contain a further element which contains the description of the current etymon.

Table 26 — Cognate class

LMF class	LBX construct
/Cognate/	<Cognate>

EXAMPLE

```
<CognateSet>
  <Cognate>
    <Lang>Chalcatongo Mixtex</Lang>
    <geoUsage>San Miguel El Grande</geoUsage>
    <Form>
      <FormRep xml:lang="mig" notation="trans-macaulay-mig">šini</FormRep>
    </Form>
    <Bibl>
```

```

        <Xref target="#Macaulay-ChalcatongoMixtec-1996">(Macaulay, 1996)</Xref>
      </Bibl>
    </Cognate>
    <Cognate>
      <Lang>Ayuta Mixtex</Lang>
      <Form>
        <FormRep xml:lang="miy" notation="trans-hill-1990-miy">shihi</FormRep>
      </Form>
      <Bibl>
        <Xref target="#Hills-AyutlaMixtec-1990">(Hills, 1990)</Xref>
      </Bibl>
    </Cognate>
    <Cognate>
      <Lang>San Martin Durazos Mixtex</Lang>
      <Form>
        <FormRep xml:lang="smd" notation="ipa">ʃiɲi</FormRep>
      </Form>
      <Bibl>
        <Xref target="#Padgett-2017">(Padgett, 2017)</Xref>
      </Bibl>
    </Cognate>
    <Bibl>
      <Xref target="#Bowers-Romary-2018">(Bowers & Romary, 2018)</Xref>
    </Bibl>
  </CognateSet>

```

9 Additional mechanisms

9.1 Overview

LBX provides additional features which do not correspond to classes or mechanisms introduced in the various models standardized in other parts of the ISO 24613 series. This clause introduces features that can be needed for a more precise representation of lexical content and meta-data in digitized dictionaries.

9.2 XML feature structure implementation

LBX provides a W3C feature structure implementation to support a more accurate representation of language specific linguistic features. Although ISO 24613-1:2019, 5.3.9, references the use of feature structures, there is no guidance on the specific mechanisms for implementation. This document describes a W3C XML schema implementation for feature structure representation (see [Annex B](#) for a more complete description). In addition, LBX provides a basic XML schema that does not include feature structures.

9.3 Representing various labels with <LBL>

The <LBL> element shall be used to encode any kind of labels used in printed dictionaries to express meta-lexicographic relations in a dictionary entry. Although LBX was designed to support large-scale electronic dictionaries, this metadata can be useful for developers and for use in results displays.

9.4 Providing rendering information with the @rend attribute

The @rend attribute can be used on all the relevant elements to provide information about the typographical features associated with the text they contain.

Annex A (informative)

LBX data category selection

A.1 Attributes allocated to the <LexicalResource>, <Lexicon> and <Entry> elements

See [Tables A.1](#) to [A.4](#).

Table A.1 — Attributes allocated to the <LexicalResource> element

Attribute	Type	Values
@languageResourceID or @xml:id	xs:ID	open (as URI)

Table A.2 — Attributes allocated to the <Lexicon> element

Attribute	Type	Values
@lexiconID or @xml:id	xs:ID	open (as URI)
@lexiconType	xs:string	closed
@sourceLanguage	xs:string	open
@targetLanguage	xs:string	open

Table A.3 — Attributes allocated to the <Entry> element

Attribute	Type	Values
@entryID or @xml:id	xs:ID	open (as URI)
@lexiconID or @xml:id	xs:ID	open (as URI)

Table A.4 — Enumeration values allocated to the Lexicon attributes — @lexiconType values

Values
bilingualDictionary
multilingualDictionary

A.2 Attributes allocated to forms

A.2.1 General

See [Tables A.5](#) to [A.8](#).

Table A.5 — Attributes allocated to the <Lemma> and <WordForm> elements

Attribute	Type	Open/Closed
@register	xs:string	open
@formType	xs:string	enumeration
@formStructure	xs:string	enumeration
@partType	xs:string	enumeration
@homonymID	xs:ID	open (as URI)
@senseRef	xs:string	open

Table A.6 — Attributes allocated to the <Stem> element

Attribute	Type	Values
@xml:lang	xs:string	BCP 47
@stemName	xs:string	open
@stemType	xs:string	enumeration

Table A.7 — Attributes allocated to the <WordPart> element

Attribute	Type	Values
partType	xs:string	enumeration

Table A.8 — Attributes allocated to the <RelForm> element

Attribute	Type	Values
@register	xs:string	open
@relType	xs:string	enumeration
@formStructure	xs:string	enumeration

A.2.2 Values for form related attributes

A.2.2.1 @register values

@register is an open data category.

A.2.2.2 @formType values

See [Table A.9](#). The @formType describes the basic characteristics of a <Lemma> or <WordForm> element. The values for the <Lemma> and <WordForm> elements are often contrasted. For example, if a <Lemma> element has a @formType attribute set to "abbreviation", a corresponding <WordForm> element might have a @formType attribute set to "fullForm".

Table A.9 — @formType values

Enumeration values
abbreviation
acronym
contraction
fullForm
inflection
variant

A.2.2.3 @formStructure values

See [Table A.10](#). @formStructure describes the component structure of a <Lemma>, <WordForm>, or <RelForm> element.

Table A.10 — @formStructure values

Enumeration values
word
MWE
particle
idiom

A.2.2.4 @partType values

See [Table A.11](#). When describing a <Lemma> or <WordForm> element, the @partType data category represents a subword particle that acts as a lemma. In this usage, the @formType and @partType data categories are mutually exclusive. When describing a <WordPart> element, the @partType specifies the type of <WordPart> that is a component of a <Lemma> element.

Table A.11 — @partType values

Enumeration values
prefix
circumfix
infix

A.3 Attributes allocated to the representation classes

A.3.1 Attributes

See [Table A.12](#). The following data categories are associated with the OrthographicRepresentation constructs, especially when used to support large scale multilingual databases. In the case of small data collections and simpler data structures, such as a single bilingual lexicon, these data categories can be managed as metadata in the <GlobalInformation> or <LexiconInformation> elements in order to provide a global scope.

Table A.12 — Attributes

Attribute	Type	Value
@xml:lang	xs:string	BCP 47
@script	xs:string	ISO 15924
@notation	xs:string	open
@geoUsage	xs:string	open
@variant	xs:string	open
@representationType	xs:string	enumeration

A.3.2 @representationType values

See [Table A.13](#).

Table A.13 — @representationType values

Enumeration values
canonicalForm
transliteration
transcription
romanization
syllabification
pronunciation

A.3.3 @notation

@notation is an open set of orthographic names for specific orthographic conventions, such as:

- **voweledArabic** Arabic script containing vowel points;
- **Pinyin Romanization** the PRC standard for Romanization;
- **Russian BGN** a Russian Romanization maintained by the US Board of Geographic Names.

The @notation will often explicitly or implicitly describe a @representationType.

A.4 Attributes allocated to the <Sense> related elements

A.4.1 Attributes allocated to the <Sense> element

See [Table A.14](#). In LBX, the <Entry> element should contain a @senseNR and/or @senseID attribute. The use of the @senseID is a best practice, in which case the @senseNR is typically not needed.

Table A.14 — Attributes allocated to the <Sense> element

Attribute	Type	Values
@senseNR	xs:string	open
@senseID	xs:string	open

A.4.2 Attributes allocated to the <Translation> element

See [Table A.15](#). In LBX, the translation text is typically in the canonical orthography for the language.

Table A.15 — Attributes allocated to the <Translation> element

Attribute	Type	Values
@xml:lang	xs:string	BCP 47
@script	xs:string	ISO 15924
@notation	xs:string	open

A.4.3 Attributes allocated to the <Example> element

See [Table A.16](#). In LBX, the example text might be in a dialectical or non-standard orthography.

Table A.16 — Attributes allocated to the <Example> element

Attribute	Type	Values
@xml:lang	xs:string	BCP 47
@script	xs:string	ISO 15924
@notation	xs:string	open

STANDARDSISO.COM : Click to view the full PDF of ISO 24613-5:2022

Annex B (informative)

LBX feature structure implementation

B.1 General

An example of an updated LBX XML schema in accordance with this document is available from Reference [7].

B.2 Target namespaces

In addition to the default LBX namespace (see [Clause 4](#)), LBX provides target namespaces to support language specific features, using unique namespace prefixes to define one or more target namespaces. The namespace convention used in LBX consists of the relevant ISO 639-3 three-character code + /ld/ (for language data). The following example shows a namespace declaration for vocabularies unique to Arabic language metadata.

EXAMPLE

```
xmlns:arald="http://www.arald.com/ArabicLanguageData"
```

B.3 Choice groups and namespaces

A choice group in XSD schemas allows a choice of elements to be made in the structure of an instance document (i.e. lexicon). When used in conjunction with namespaces, the choice can be used to determine the language of interest for the feature structures. Elements in the choice group reference a URI consisting of a relevant language namespace using the convention described above followed by the name of a target element. The below example shows a choice to be made among Arabic, Russian, French, Spanish, and Zulu grammatical features.

EXAMPLE 1

```
<xs:group name="GramFeats">
  <xs:choice>
    <xs:element ref="arald:GramFeats"/>
    <xs:element ref="rusld:GramFeats"/>
    <xs:element ref="frald:GramFeats"/>
    <xs:element ref="espld:GramFeats"/>
    <xs:element ref="zullld:GramFeats"/>
  </xs:choice>
</xs:group>
```

The choice group for grammatical features is then referenced as a child of the appropriate element in the LBX schema. In the following example, the complex type includes a reference to the GramFeats choice group that is applicable for all declared languages.

EXAMPLE 2

```
<xs:complexType name="WordFormType">
  <xs:sequence>
    <xs:group ref="GramFeats" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element ref="FormRep" minOccurs="0" maxOccurs="unbounded"/>
  </xs:sequence>
  <xs:attribute ref="formType" use="optional"/>
  <xs:attribute ref="formStructure"/>
</xs:complexType>
```

```
<xs:element name="WordForm" type="WordFormType"/>
<xs:element name="Lemma" type="WordFormType"/>
```

A lexicon (XML document) can then reference any of the “GramFeats” elements declared in the choice group by referencing the relevant namespace. The following example illustrates the use of this feature structure mechanism for a Spanish language lexicon.

EXAMPLE 3

```
<Entry>
  <Lemma>
    <espld:GramFeats>
      <espld:POS>noun</espld:POS>
      <espld:Gender>masculine</espld:Gender>
      <espld:Number>plural</espld:Number>
    </espld:GramFeats>
    [one or more <FormRep> elements]
  </Lemma>
  <Sense/>
</Entry>
```

For comparison, an Arabic language feature structure attribute-value pair <arald:Number>dual</arald:Number> is not relevant for Spanish, since Spanish does not have a “dual” grammatical number.

B.4 Implementing features and values

LBX manages grammatical features by means of XSD documents specific to each language. For example, grammatical features for Arabic are managed in a document with the namespace declared as follows:

```
xmlns:arald:="http://www.arald.com/ArabicLanguageDocument"
```

A <complexType> element references the grammatical features relevant for the language. The following example shows a partial set of grammatical features for Arabic (for simplicity, cardinality is not shown).

EXAMPLE 1

```
<xs:complexType name="GrammaticalFeaturesType">
  <xs:sequence>
    <xs:element ref="Person"/>
    <xs:element ref="Gender"/>
    <xs:element ref="Number"/>
  </xs:sequence>
</xs:complexType>
```

The <GramFeats> element in the “arald” namespace is then assigned to the above type:

```
<xs:element name="GramFeats" type="GrammaticalFeaturesType"/>
```

The allowable values for a specific grammatical feature are then defined using <xsd:simpleType> and the element representing that feature is assigned to the appropriate <xsd:simpleType>, such as in the following example.

EXAMPLE 2

```
<xs:simpleType name="GrammaticalGenderType">
  <xs:restriction base="xs:NMTOKEN">
    <xs:enumeration value="masculine"/>
    <xs:enumeration value="feminine"/>
  </xs:restriction>
</xs:simpleType>
<xs:element name="Gender" type="GrammaticalGenderType"/>
```

B.5 Feature value management

Values for elements contained in the grammatical features type, such as “grammaticalNumberType”, can be defined through XSD <simpleType> elements which are then declared for the appropriate grammatical feature element, as in the following example.

EXAMPLE

```
<xs:element name="Number" type="GrammaticalNumberType">
<xs:simpleType name="GrammaticalNumberType">
  <xs:restriction base="xs:string">
    <xs:enumeration value="singular"/>
    <xs:enumeration value="plural"/>
    <xs:enumeration value="dual"/>
  </xs:restriction>
</xs:simpleType>
</xs:element>
```

LBX grammatical features provide a very powerful tool for ensuring data consistency and accuracy for each language. The use of these features also provides database design advantages, for example, providing XML support for the coding of relevant picklists that support the display of lexicon search results and the editing of new lexicon entries.