

Lexical Markup Framework

In 2008 ISO TC 37 published the Lexical Markup Framework (LMF; ISO 24613:2008), which provides an object model, specified in Unified Modeling Language (UML) class diagrams, for lexical resources. The object model is a series of extensions of a LMF core model.

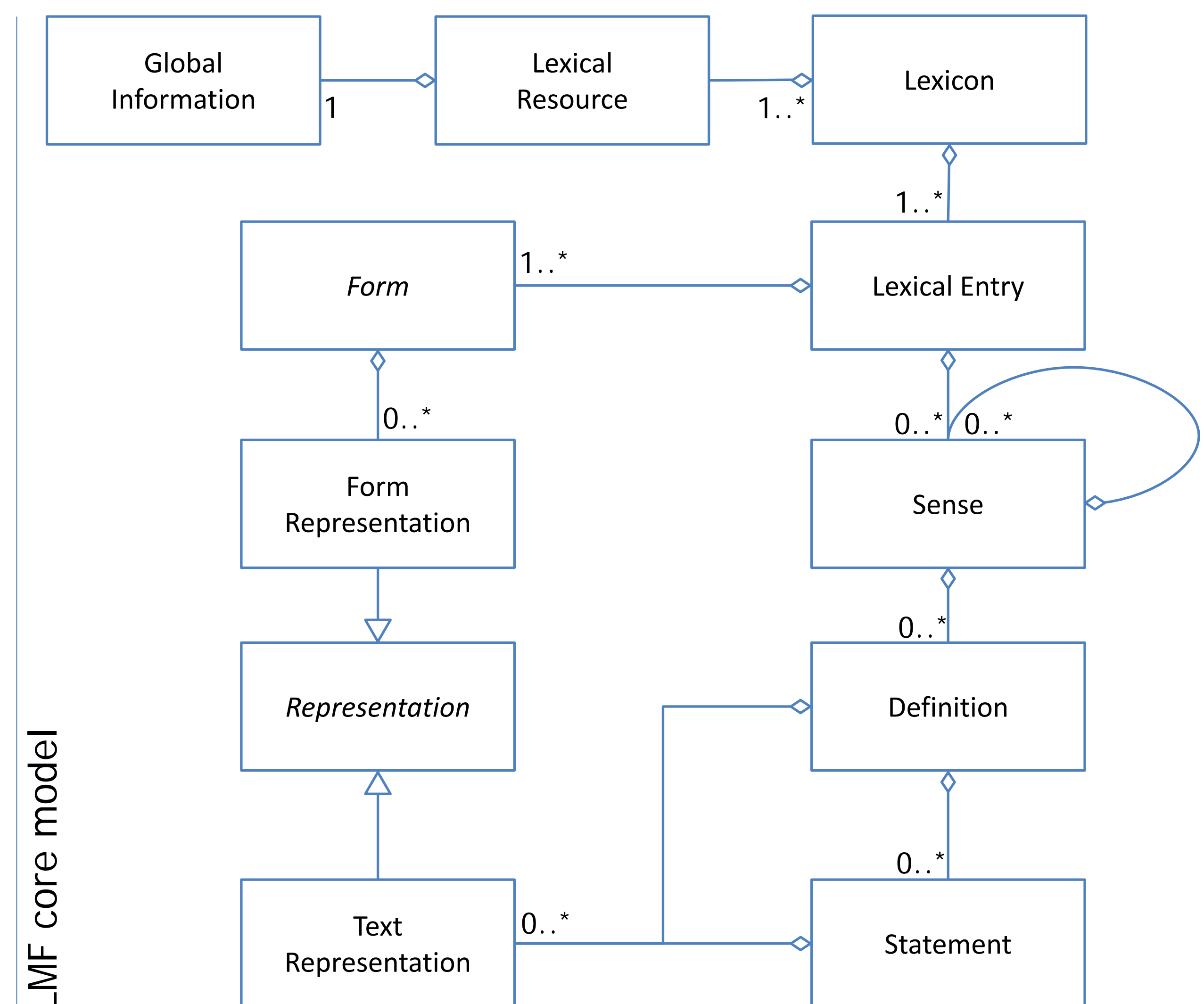
Extensions

1. Morphology
2. Machine readable dictionary
3. NLP syntax
4. NLP semantics
5. NLP multilingual notations
6. NLP morphological patterns
7. NLP multiword expression patterns
8. Constraint expression

“Additional extensions may be developed over time.”

Feature Structures

Adorn the classes with features selected from the ISOcat Data Category Registry (www.isocat.org). As ISOcat is an open registry and LMF allows for extensions anyone can tweak the LMF model to the needs of their lexical resource, as long as they stay compliant with the core model.



Interoperability?

LMF provides good practices in the form of patterns and terminology for common data modeling needs for lexical resources. However, the (XML) serialization of the model in Annex R is informative, and is explicitly left open: “A user can decide to define another DTD or schema to implement LMF.”

RELISH LMF

<http://tla.mpi.nl/relish/lmf/>

In the DFG/NEH Rendering Endangered Languages Lexicons Interoperable through Standards Harmonization (RELISH) project LMF played the role of a pivot format for lexical resources. However, the informative LMF serialization is too weak, so an own serialization was defined which brings all the power of LMF to the data interoperability level.

Two cooperating modern XML schema/validation languages

RELISH LMF uses two modern XML schema languages: Relax NG (ISO/IEC 19757-2:2008) and Schematron (ISO/IEC 19757-3:2006). These allow a more modular approach, i.e., one can actually select the needed LMF extensions and define ones own extensions. Also the use of Schematron rules as an additional validation layer allows to express more of the restrictions present in the standardized LMF UML class diagrams, e.g., every Lexical Entry should have a Form.

Rich Feature Structures

The standard mentions that the rich ISO/TEI Feature Structure Representation (ISO 24610-1:2006) can be used as an alternative to simple feature structures allowed in the informative DTD, i.e., with nesting of feature structures and typing of feature values. Additionally RELISH LMF allows to include a ISO/TEI Feature System Declaration (ISO 24610-2:2011), paving the way for future validation of feature structures.

Persistent references to ISOcat Data Categories

In the informative DTD the link to ISOcat Data Categories is underspecified, i.e., incorrectly an unique Data Category identifier is assumed and this can lead to ambiguity and shifting semantics. RELISH LMF uses the ISOcat Data Category Reference vocabulary to embed the ISOcat Persistent Identifiers in a feature structure/system.

```
<grammar xmlns="http://relaxng.org/ns/structure/1.0"
  datatypeLibrary="http://www.w3.org/2001/XMLSchema-datatypes"
  xmlns:lmf="http://www.lexicalmarkupframework.org/"
  <!-- 0. always -->
  <include href="path-to-relish-lmf/RELISH-LMF-common.rng"/>
  <include href="path-to-relish-lmf/RELISH-LMF-core.rng"/>
  <!-- 1. choose a feature structure representation -->
  <include href="path-to-relish-lmf/RELISH-LMF-fs-tei.rng"/>
  <!-- 2. optionally: select LMF extensions -->
  <include href="path-to-relish-lmf/RELISH-LMF-morphology.rng"/>
  <!-- 3. optionally: add your own extensions -->
  <!-- Define an User element (using example.com as the XML namespace URL) -->
  <define name="myexample.User">
    <element name="User" ns="http://example.com">
      <ref name="myexample.User.attributes"/>
      <ref name="myexample.User.content"/>
    </element>
  </define>
  <define name="myexample.User.attributes">
    <ref name="relish.lmf.common.attributes"/>
  </define>
  <define name="myexample.User.content">
    <interleave>
      <ref name="relish.lmf.fs"/>
    </interleave>
  </define>
  <!-- Link User to Lexical Resource -->
  <define name="relish.lmf.LexicalResource.content" combine="interleave">
    <interleave>
      <zeroOrMore>
        <ref name="myexample.User"/>
      </zeroOrMore>
    </interleave>
  </define>
</grammar>
```

```
<?xml-model href="path-to-your-lexicon/schema.rng" type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0?>
<?xml-model href="path-to-your-lexicon/schema.rng" type="application/xml" schematypens="http://purl.oclc.org/dsdl/schematron?>
<LexicalResource xmlns="http://www.lexicalmarkupframework.org/" lmfVersion="ISO 24613:2008"
  xmlns:dcr="http://www.isocat.org/ns/dcr" xmlns:ex="http://example.com" xmlns:tei="http://www.tei-c.org/ns/1.0">
  <ex:User xml:id="jd">
    <tei:f name="first name" dcr:datcat="http://www.isocat.org/datcat/DC-4194">
      <tei:string>John</tei:string>
    </tei:f>
    <tei:f name="last name" dcr:datcat="http://www.isocat.org/datcat/DC-4195">
      <tei:string>Doe</tei:string>
    </tei:f>
  </ex:User>
  <GlobalInformation>
    <tei:f name="languageCoding" dcr:datcat="http://www.isocat.org/datcat/DC-2008">
      <tei:symbol value="ISO 639-3"/>
    </tei:f>
  </GlobalInformation>
  <Lexicon xml:lang="en">
    <tei:f name="language" dcr:datcat="http://www.isocat.org/datcat/DC-1969">
      <tei:symbol value="eng"/>
    </tei:f>
    <LexicalEntry xml:id="le0001">
      <tei:f name="partOfSpeech" dcr:datcat="http://www.isocat.org/datcat/DC-1345">
        <tei:symbol value="commonNoun" dcr:valueDatcat="http://www.isocat.org/datcat/DC-1256"/>
      </tei:f>
      <Lemma type="Form">
        <tei:f name="writtenForm" dcr:datcat="http://www.isocat.org/datcat/DC-1836">
          <tei:string>clergyman</tei:string>
        </tei:f>
      </Lemma>
    </LexicalEntry>
  </Lexicon>
</LexicalResource>
```

RELISH LMF schema for an lexical resource

RELISH LMF lexical resource