LMF Revisited

Anas Fahad Khan[1], Francesca Frontini[1], Laurent Romary[2]

[1]Istituto di Linguistica Computazionale "A . Zampolli", CNR, Italy
[2]INRIA / Paris, France
[1]*{fahad.khan, francesca.frontini}@ilc.cnr.it*
[2]*laurent.romary@inria.fr*

Introduction

- Shared standards are essential for ensuring the interoperability and re-usability of language resources.
- Especially important for computational lexicons contatining highly structured information.
- The influential **Lexical Markup Framework (LMF)**, first published in 2008 by the International Standards Organization (ISO) as **ISO standard 24613:2008** was intended as a "standardized framework for the construction of computational lexicons" " (Francopoulo, 2013).
- In our poster we take a brief look at the original LMF and provide an update on the new version of LMF!

The Original LMF

- The original LMF specifications were intended to cover as wide a range of lexicon-like resources as possible.
- They made specific provision for both NLP dictionaries and *Machine Readable Dictionaries*, as well as several other categories of lexicon or lexico-semantic resource, such as for example *bilingual* and *multilingual lexicons* along with Wordnets.
- The original LMF specifications consisted of a core model together with the following extension packages: **Machine-readable Dictionary**, **Morphology, Syntax and Semantics**, **Multilingual Notation**, **Multiword Expression Pattern**, **Constraint Expression**.

The NEW LMF

- The new LMF is a multi-part standard consisting of six separate modules, each published as a separate ISO standard, with further extensions planned to come.
- It has been decoupled from any single serialisation format, although two recommended serialisations of the meta-model constitute the fourth and fifth parts of the standard (these are TEI and LBX respectively).
- The new emphasis on abstraction and modularisation has also led to a series of major simplifications affecting nearly every part of the new version of the LMF meta-model.

The New LMF

- **ISO 24613-1:2019 Language resource management — Lexical markup framework (LMF) — Part 1: Core model**: *This module defines the basic classes required to model a baseline lexicon and is a pre-requisite for the use of the other classes. Status: Published in 2019 it is now being further revised to make it easier to use.*
- **ISO 24613-2:2020 Language resource management — Lexical markup framework (LMF) — Part 2: Machine-readable dictionary (MRD) model:** *Contains components providing deeper specification of lexical description encapsulated within the core model. Status: Published in 2020.*
- **ISO 24613-3:2021 Language resource management — Lexical markup framework (LMF) — Part 3: Etymological extension**: *A completely new addition to the LMF meta-model covering etymological and diachronic information. This part makes etymologies, etymological links and etymons first class citizens. Status: Published in 2021.*

The New LMF

- **ISO 24613-4:2021 Language resource management — Lexical markup framework (LMF) — Part 4: TEI serialization:** A *TEI serialisation of the other parts of the model which aims to make both TEI and LMF fully compatible and which leverages the knowledge and makes use of the established practices of the TEI community in dealing with lexiocraphic resources. Status: Published in 2021.*
- **ISO 24613-5:2022 Language resource management — Lexical markup framework (LMF) — Part 5: Lexical base exchange (LBX) serialization:** *Another XML serialisation. Status: Published in 2022.*
- **ISO/CD 24613-6 Language resource management — Lexical markup framework (LMF) — Part 6: Syntax and Semantics:** *An update to the Syntax and Semantics parts of the original standard. Status: A candidate for an ISO Draft International Standard (DIS) ballot.*