

Toward Best Practice Guidelines for Encoding Legacy Dictionaries

An ENeL-DARIAH-PARTHENOS Expert Workshop

Preußische Staatsbibliothek (Prussian State Library)
Unter den Linden 8
10117 Berlin

November 17-19, 2016.

The workshop will consist of a sequence of parallel sessions in small groups, followed by plenary discussions. For each topic addressed during parallel sessions, the objective is to identify elements of consensus and identify further or deeper discussion points that we need to address in the next round.

We do not aim to specify here a mandatory format for the variety of dictionary content that we deal with, but define a baseline encoding (TEI-Lex-0) similar to TEI Simple against which existing dictionaries can be compared and which could serve as a target transformation format for generic querying or visualization tools.

To initiate the dynamics, we suggest the following first series of topics:

- *organizing lexical information in entries*: how to recommend usage of <entry>, <entryFree>, <superEntry> (as well as <dicScrap>); what kind of precise lexical situation should lead to the choice of one or the other; how will this affect a digitization workflow?
- *representing form*: which precise recommendations can be made concerning <form>, its usage for lemma or inflected forms, the articulation with grammatical information (<gramGrp>) or specific form surrogates such as roots;
- *organizing sense*: what, if anything, should be left unencoded? should <sense> be mandatory? can we identify model constructs for the first level of sense children (systematicity of <usg> types, guidelines for <cit>, ...);
- *towards a deeper encoded <etym>*: following the recent proposals found in <https://hal.inria.fr/hal-01296498>, shall we define a formal variant to <etym>?

Further discussion topics could deal with compound lexical items, unifying intra- and inter-lexical references (<xref>, <oRef>/<pRef>, <mentioned>, <term>), providing guidance concerning related entries, recommendations concerning the use of <gloss> vs. <cit type="translation">, dealing with punctuations and metamarks, @xml:lang on <entry>, generalisation of @notation etc.

We will jointly decide on the topics to be covered by each subsequent breakout session.

Requirements

We do not expect participants to prepare formal presentations, but we do expect everybody to bring their own dictionary examples which illustrate issues that could/should be addressed by our Guidelines.

Schedule

Thursday, November 17th	
13:00 - 13:30	Introductions
13:30 - 13:45	<i>Lexical Resources & TEI: Why We Need More, Not Fewer, Constraints</i> (Toma)
13:45 - 14:00	<i>TEI & LMF: It's Happening</i> (Laurent)
14:00 - 15:00	Breakout Sessions + Reports
15:00 - 15:30	Coffee Break
15:30 - 17:30	Breakout Sessions + Reports
Friday, November 18th	
09:00 - 11:00	Breakout Sessions + Reports
11:00 - 11:30	Coffee Break
11:30 - 13:00	Breakout Sessions + Reports
13:00 - 14:00	Lunch
14:00 - 16:00	Breakout Sessions + Reports
16:00 - 16:30	Coffee Break
16:30 - 18:00	Breakout Sessions + Reports
Saturday, November 19th	
09:00 - 11:00	Breakout Sessions + Reports
11:00 - 11:30	Coffee Break
11:30 - 13:00	Wrap-up: <i>Toward the Lexical Data Seal of Approval</i> (discussion)