Towards a Repository of Senses for Use in TEI encoded Dictionaries

Thierry Declerck, *DFKI GmbH, Language Technology Lab, Saarbrücken, Germany*
Karlheinz Mörth, *ACDH @ Austrian Academy of Sciences, Vienna, Austria*

declerck@dfki.de, Karlheinz.Moerth@oeaw.ac.at

senses

Abstract entry with 2 senses

```
<entry>
  <sense n="1"/>
  <sense n="2"/>
</entry>
```

The TEI approach to encoding of senses is described in chapter 9 "Dictionaries" of the TEI Guidelines (TEI Consortium 2016, http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf), which is dedicated to the representation of lexical resources:

"**<sense>** groups together all information relating to one word sense in a dictionary entry, for example definitions, examples, and translation equivalents." (TEI Consortium 2016: 278).

A TEI sense can include <*usg*>, <*def*>, <*cit*> elements, whereas the <*cit*> element "contains a quotation from some other document, together with a bibliographic reference to its source. In a dictionary it may contain an example text with at least one occurrence of the word form, used in the sense being described, or a translation of the headword, or an example." (TEI Consortium 2016: 280).

Abstract entry with two homographs, the first with two senses and the second with three sense

```
<entry>
  <hom n="1">
    <sense n="1">
      <!-- ... -->
    </sense>

    <sense n="2">
      <!-- ... -->
    </sense>
  </hom>

  ...
```

```
  ...

  <hom n="2">
    <sense n="1">
      <sense n="a">
        <!-- ... -->
      </sense>
      <sense n="b">
        <!-- ... -->
      </sense>
    </sense>
    <sense n="2">
      <!-- ... -->
    </sense>
    <sense n="3">
      <!-- ... -->
    </sense>
  </hom>
</entry>
```

No defined restrictions as how to codify the content of the sense element, and all possible string characters seem to be allowed.
Traditionally lumping together several related senses in one quote (e.g. <cit><quote>door, gate</quote></cit>) is quite common too.
Not optimal to rely on string matching for stating a relation between senses included in different entries in different dictionaries.

Towards a repository of formally well defined senses for improved

- Update mechanisms
- Maintainability
- Accessability

of senses, also accross dictionaries

<sense> pointing to a repository of data categories, via the <ptr> element

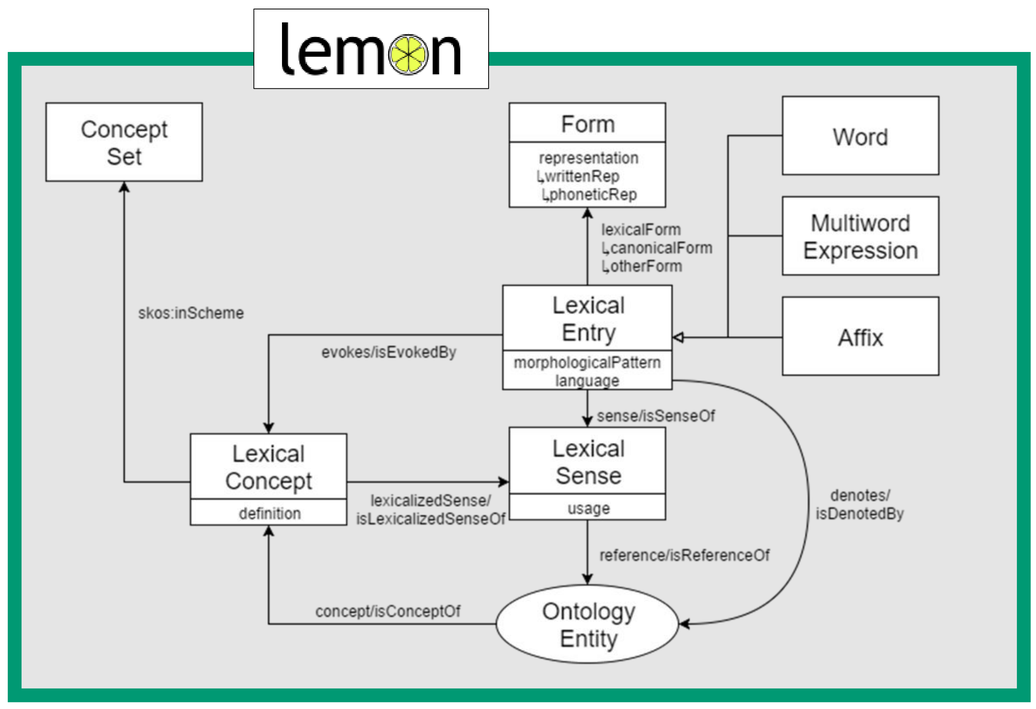
Describing the entry-sense relation on the basis of LMF, *lemon* approach for lexicon modelling

BASE

A native and light-weight XML database management system

Technologies
- xPath,
- xQuery
- APIs: RESTful API, WebDAV, etc.

Many supported formats and Integration in Virtual Research Environment,
like the Viennese Lexicographic Editor



The core model of OntoLex. Figure created by John P. McCrae for the W3C OntoLex Community Group