

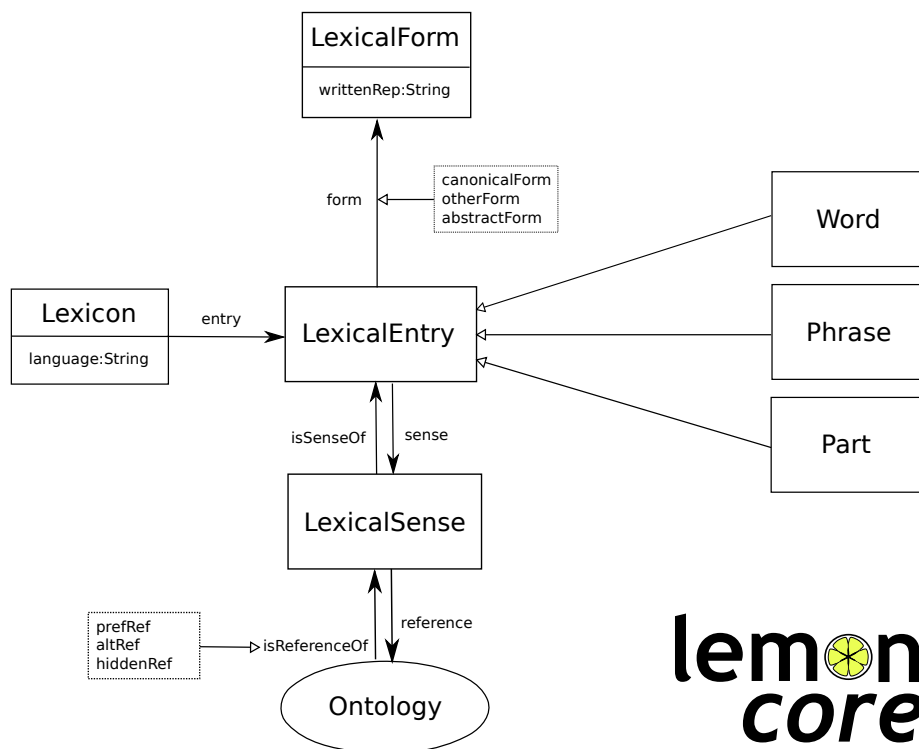
lemon in 5 minutes

John McCrae¹

¹ CITEC, Universität Bielefeld



lemon in 5 minutes



lemon
core

lemon is an RDF model for representing lexical information relative to ontologies. We shall assume the reader is familiar with the concepts of RDF and understands the Turtle syntax for RDF.

The lemon model consists of a core path defined as:

Ontology Entity ↔ Lexical Sense ↔ Lexical Entry → Lexical Form → (Written) Representation

For example the lexical entry for a simple word such as “tuberculosis” would be as follows:

```

:tuberculosis lemon:canonicalForm [
  lemon:writtenRep "tuberculosis"@en ] ;
lemon:sense [
  lemon:reference <http://dbpedia.org/resource/Tuberculosis> ] .
  
```

Example 1

Lexica are monolingual

Each lexicon collects all entries for a given language, as such language is normally marked on the lexicon object. e.g.,

```

:lexicon lemon:language "en" ;
  lemon:topic <http://dbpedia.org/resource/Disease> ;
  lemon:entry :tuberculosis , :consumption , :hepatitis .
  
```

Example 2

Note that forms are also by convention marked with a language tag, this tag may be more specific (e.g., “en-GB” for “UK English” and “en-fonipa” for “English IPA” are allowed in the lexicon above)

An entry has one morphosyntax

Each entry in a lexicon is assumed to have the same part-of-speech class and if it is multi-word the same parse tree. e.g.,

```
# Derived forms are separate entries
:garlic lemon:canonicalForm [
  lemon:writtenRep "garlic"@en ] .

:garlicky lemon:canonicalForm [
  lemon:writtenRep "garlicky"@en ] .

# Term variations are separate entries
:garlic_clove lemon:canonicalForm [
  lemon:writtenRep "garlic clove"@en ] ;
lemon:sense [
  lemon:reference <http://dbpedia.org/resource/Garlic> ] .

:clove_of_garlic lemon:canonicalForm [
  lemon:writtenRep "clove of garlic"@en ] ;
lemon:sense [
  lemon:reference <http://dbpedia.org/resource/Garlic> ] .
```

Example 3

A form is spoken one way

A form may have multiple representations, for example to capture phonetics and spelling variation, however it must always be pronounced the same way. A different form is used for each inflection variant:

```
:color lemon:canonicalForm [
  lemon:writtenRep "color"@en-US ;
  lemon:writtenRep "colour"@en-GB ;
  lemon:representation "'kʌl.ə(ɹ)"@en-fonipa ] ;
lemon:otherForm [
  lemon:writtenRep "colors"@en-US ;
  lemon:writtenRep "colours"@en-GB ;
  lemon:representation "'kʌl.ə:s"@en-fonipa ] ;
lemon:sense [
  lemon:reference <http://dbpedia.org/resource/Color> ] .
```

Example 4

There is one sense for each entry-entity pair

The sense object represents a mapping between a lexical entry and an ontology entity. As such, no two lexical entries may have the same sense object, and the sense object should have only one reference, except in the case when the ontology entities are known to be equivalent by a link such as `owl:sameAs`.

```
:tuberculosis lemon:canonicalForm [
  lemon:writtenRep "tuberculosis"@en ] ;
lemon:sense [
  lemon:reference <http://dbpedia.org/resource/Tuberculosis> ] .

# Consumption is an alternative term for tuberculosis, only antiquated.
# The word also has a modern meaning in economics.
:consumption lemon:canonicalForm [
  lemon:writtenRep "consumption"@en ] ;
```

```

lemon:sense [
  lemon:reference <http://dbpedia.org/resource/Tuberculosis> ;
  lemon:context :antiquated ] ;
lemon:sense [
  lemon:reference <http://dbpedia.org/resource/Consumption_(economics)> ] .

# OK as we know that
# <http://dbpedia.org/resource/Garlic> owl:sameAs
#   <http://rdf.freebase.com/ns/m/0dbrl> .
:garlic lemon:canonicalForm [
  lemon:writtenRep "garlic"@en ] ;
lemon:sense [
  lemon:reference <http://dbpedia.org/resource/Garlic> ,
                 <http://rdf.freebase.com/ns/m/0dbrl> ] .

```

Example 5