

WordNet LMF and JSON

John P. McCrae

Insight Centre for Data Analytics at National
University of Ireland, Galway

WordNet LMF and JSON

- Two formats for creating WordNets
- WordNet LMF is 'LMF-like' XML format
- WordNet JSON is JSON-LD based representation
 - This is RDF/Linked Data
 - Can be further converted to RDF/XML, Turtle, etc.
- These formats are isomorphic
- There are converters:
 - <http://github.com/globalwordnet/schemas/tree/master/converter>
 - They double (for the moment) as validators
- Either form can be used for submission to CILI
- Stable final version coming soon (i.e., this week)

WordNet-LMF (Example)

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE LexicalResource SYSTEM "http://globalwordnet.github.io/schemas/WN-LMF.dtd">
<LexicalResource>
  <Lexicon label="Princeton WordNet" language="en">
    <LexicalEntry id="w1">
      <Lemma writtenForm="entity" partOfSpeech="n"/>
      <Sense id="pwn-100001740-n-1" synset="pwn-100001740-n"/>
    </LexicalEntry>
    <Synset id="pwn-100001740-n" ili="i35545">
      <Definition gloss="that which is perceived or known or inferred to have its own distinct existence (living or nonliving)"
iliDef=""/>
      <SynsetRelation relType="hyponym" target="pwn-104431553-n"/>
    </Synset>
    <Meta publisher="Global Wordnet Association" description="Example lexicon" />
  </Lexicon>
</LexicalResource/>
```

WordNet-LMF

- Use shortest ISO 639 code (e.g., “en”, “tlh”, “vo-rigik”) as BCP 47
- Give an id to every element that clearly identifies the source, version and element
 - They must be unique in your file (of course)
- Parts of speech: n,v,a,r,s,p,u
- Please add the ILI code or “in” to propose a novel synset
- iliDef is not necessary (unless proposing a novel synset) but must be in English
- relationType must be one from a fixed list
 - Other relation types are not supported
- Meta can go in many places

WordNet-JSON (Linked Data)

```
{ "@context": [ "http://globalwordnet.github.io/schemas/wn-json-context.json", { "@language": "en" } ],
  "@id": "pwn30",
  "@type": "lemon:Lexicon",
  "label": "Princeton WordNet",
  "language": "en",
  "publisher": "Princeton University",
  "rights": "wordnetlicense:",
  "entry": [{
    "@id": "w1",
    "lemma": { "writtenForm": "entity" },
    "partOfSpeech": "wn:noun",
    "sense": [{
      "@id": "pwn-100001740-n-1",
      "synset": {
        "@id": "pwn-100001740-n",
        "iliRef": "ili:i35545",
        "definition": {
          "gloss": "that which is perceived or known or inferred to have its own distinct existence (living or nonliving)"
        },
        "hyponym": ["pwn-104431553-n", "pwn-100002137-n", "pwn-100001930-n"]
      }
    }
  ]
}
```

WordNet-JSON

- Please try to keep to the structure
 - i.e., Don't use generic RDF to JSON-LD converters
- Synsets are nested under senses (!)
 - Second occurrence need only give “@id”
- JSON is flexible (new properties can be added)
- Based on following vocabularies
 - Princeton WordNet RDF
 - *lemon* (maybe OntoLex?)
 - Dublin Core
 - SKOS
 - RDFS