

Using OntoLex-Lemon for Representing and Interlinking German Multiword Expressions in OdeNet and MMORPH

Thierry Declerck

¹ German Research Center
for Artificial Intelligence
Saarbrücken, Germany

² Austrian Centre
for Digital Humanities
Vienna, Austria

thierry.declerck@dfki.de

Melanie Siegel

Darmstadt University
of Applied Sciences
Darmstadt, Germany

melanie.siegel@h-da.de

Stefanie Racioppa

German Research Center
for Artificial Intelligence
MLT Lab

Saarbrücken, Germany

stefania.racioppa@dfki.de

Abstract

We describe work consisting in porting two large German lexical resources into the OntoLex-Lemon model in order to establish complementary interlinkings between them. One resource is OdeNet (Open German WordNet) and the other is a further development of the German version of the MMORPH morphological analyzer. We show how the Multiword Expressions (MWEs) contained in OdeNet can be morphologically specified by the use of the lexical representation and linking features of OntoLex-Lemon, which also support the formulation of restrictions in the usage of such expressions.

1 Introduction

WordNets are well-established lexical resources with a wide range of applications. For more than twenty years they have been elaborately set up and maintained by hand, especially the original Princeton WordNet of English (PWN) (Fellbaum, 1998). In recent years, there have been increasing activities in which open WordNets for different languages have been automatically extracted from other resources and enriched with lexical semantics information, building the so-called Open Multilingual WordNet (Bond and Paik, 2012). These WordNets were linked to PWN via shared synset. In this context a German lexical semantics resource with the name Open German WordNet (OdeNet)¹ is being developed with the aim to be included as the first open German WordNet into the Open Multilingual WordNet.

This paper deals with the morphological enrichment of OdeNet, with a focus on complex OdeNet entries. The first morphological resource we are

considering for this task is an updated German version of the MMORPH morphological analyzer (Petitpierre and Russell, 1995).² Besides this resource we have consulted the on-line editions of Duden and CanooNet,³ as well as entries in the German Wiktionary⁴ for manually checking a few lexical features of both OdeNet and MMORPH.

As a representation mean we have adopted OntoLex-Lemon (Cimiano et al., 2016),⁵ as this model was shown to be able to represent both classical lexicographic description (McCrae et al., 2017) and lexical semantics networks, like WordNet (McCrae et al., 2014). OntoLex-Lemon is a further development of the “LEXicon Model for Ontologies” (*lemon*).⁶ Guidelines for mapping Global WordNet formats onto *lemon*-based RDF⁷ have been published⁸ and already some WordNets have been mapped onto *lemon*, as described for example in (McCrae et al., 2014).

We follow in this work the suggestion made in (Hüning and Schlücker, 2015) to consider MWEs as being “a general term that includes phenomena with different degrees of syntactic fixedness and semantic compositionality”, allowing us to treat German compounds in a similar way as

²The German version of this analyzer has been further developed, also improving the inclusion of compounds, and the resulting OntoLex-Lemon representation of this extended resource will be made publicly available, in the Linguistic Linked Open Data cloud (see <http://linguistic-lod.org/llod-cloud>).

³See <https://www.duden.de/woerterbuch> and <http://www.canoo.net/> respectively.

⁴<https://de.wiktionary.org/wiki/Kategorie:Deutsch>.

⁵See also <https://www.w3.org/2016/05/ontolex/> for more details.

⁶See (McCrae et al., 2012)

⁷RDF stands for “Resource Description Framework”. See <https://www.w3.org/RDF/> for more details.

⁸See <https://globalwordnet.github.io/schemas/#rdf>.

¹We collected information on OdeNet from <https://github.com/hdaSprachtechnologie/odenet>.

OdeNet “phrasal entries”, so that OdeNet entries “Rotkraut” (*red kraut* or *red cabbage*) and “rote Bete” (*beetroot*) can be equally considered as MWEs, but both terms will be associated to different morphological patterns, as the German adjective “rote” in the second case is also displaying an inflectional behavior in order to be in agreement with the morphology of the noun (for example in singular genitive or in all forms of plural, and also in dependency of the preceding presence of a definite or indefinite determiner).

In the next sections, we give first background information on OdeNet and on OntoLex-Lemon. We then describe the mapping of OdeNet to OntoLex-Lemon. We continue with an introduction of MMORPH, followed by a section that describes how the use of MMORPH and OntoLex-Lemon is supporting the linking of MWEs in OdeNet to full morphological descriptions.

2 OdeNet

A candidate for representing German lexical semantics data in OntoLex could be for sure GermaNet, which is a manually well-designed WordNet resource for German (Hamp et al., 1997). GermaNet was developed over 20 years now and is very stable and precise. The problem with GermaNet is that it is not available under an open-source license. The restricted license makes GermaNet unable to be included in the aforementioned Open Multilingual WordNet. Therefore we selected OdeNet as the German lexical semantics resource we want to work with, also with the aim of publishing the resulting data set as part of the Linguistic Linked Open Data cloud.⁹

OdeNet combines two existing resources: The OpenThesaurus German synonym lexicon¹⁰ and the Open Multilingual WordNet (OMW)¹¹ English resource: the Princeton WordNet of English (PWN) (Fellbaum, 1998). Considering the integration of OpenThesaurus in OdeNet means making use of a large resource for German that is generated and updated by the crowd. A consequence of this approach is that OdeNet needs to be curated. While generally automatically generated

⁹In a next step we will also consider the resource lemonUby (Eckle-Kohler et al., 2015), which contains a lemon representation of the German version of Omega-Wiki. A dump of this resource can be downloaded at https://lemon-model.net/lexica/uby/ow_deu/.

¹⁰<https://www.openthesaurus.de/>

¹¹<http://compling.hss.ntu.edu.sg/omw/>. See also (Bond and Foster, 2013).

entries have a confidence score of “0.7”, manually curated entries get a score of “1.0”.

We downloaded the most recent version of OdeNet from its GitHub page,¹² and first analyzed its content. The resource comes in an XML format and shares its DTD with the other WordNets in the Open Multilingual WordNet initiative.¹³ Lexical entries give information about the sense of the lexeme, such as “Kernspaltung” or “Kernfission” (*nuclear fission*), both sharing the same synset:¹⁴

```
<LexicalEntry id="w1">
  <Lemma writtenForm="Kernspaltung"
    partOfSpeech="n"/>
  <Sense id="w1_1-n"
    synset="odenet-1-n"/>
</LexicalEntry>
<LexicalEntry id="w2">
  <Lemma writtenForm="Kernfission"
    partOfSpeech="n"/>
  <Sense id="w2_1-n"
    synset="odenet-1-n"/>
</LexicalEntry>
```

Lexical senses are grouped in synsets, i.e., groups of word senses with the same meaning. Hierarchical relations are introduced as synset relations:

```
<Synset id="odenet-1-n" ili="i107577"
  partOfSpeech="n" dc:description="a
  nuclear reaction in which a
  massive nucleus splits into
  smaller nuclei with the
  simultaneous release of energy">
  <SynsetRelation
    target='odenet-5437-n'
    relType='hypernym' />
</Synset>
```

An example for the curated entry “Stuhl”: (*chair*):

```
<LexicalEntry id="w224"
  confidenceScore="1.0">
  <Lemma writtenForm="Stuhl"
    partOfSpeech="n"/>
  <Sense id="w224_49-n"
    synset="odenet-49-n"/>
  <Sense id="w224_1172-n"
    synset="odenet-1172-n"/>
</LexicalEntry>
<Synset id="odenet-49-n" ili="i51746"
  partOfSpeech="n" confidenceScore="1.0">
  <Definition>
    Eine Sitzgelegenheit fuer eine Person,
    mit einer Lehne im Ruecken.
  </Definition>
  <SynsetRelation target='odenet-11251-n'
    relType='hypernym' />
  <SynsetRelation target='odenet-8518-n
```

¹²<https://github.com/hdaSprachtechnologie/odenet>.

¹³See <https://github.com/globalwordnet/schemas/blob/master/WN-LMF.dtd> for more details.

¹⁴These are automatically generated and not yet curated entries that got their synset definition from an automatic linking to PWN.

```

    relType='hyponym' />
<SynsetRelation target='odenet-20127-n'
  relType='hyponym' />
<SynsetRelation target='odenet-34983-n'
  relType='hyponym' />
<Example>
  Sie sitzt auf dem Stuhl.
</Example>
</Synset>

```

Access to the lemma information for hypernyms and hyponyms is also possible, so for the odenet-49-n synset for “Stuhl”:

```

>>> hypernyms("odenet-49-n")
odenet-11251-n:
['Sitz', 'Platz', 'Sitzplatz',
 'Sitzgelegenheit']

>>> hyponyms("odenet-49-n")
odenet-8518-n:
['Rolli', 'Krankenfahrstuhl',
 'Rollstuhl'],
odenet-20127-n:
['Lehnsessel', 'Fauteuil'],
odenet-34983-n:
['Lehnstuhl', 'Polsterstuhl',
 'Polstersessel', 'Sessel', ...]]

```

3 OntoLex-Lemon

The OntoLex-Lemon model was originally developed with the aim to provide a rich linguistic grounding for ontologies, meaning that the natural language expressions used in the description of ontology elements are equipped with an extensive linguistic description.¹⁵ This rich linguistic grounding includes the representation of morphological and syntactic properties of lexical entries as well as the syntax-semantics interface, i.e. the meaning of these lexical entries with respect to an ontology or to specialized vocabularies. The main organizing unit for those linguistic descriptions is the lexical entry, which enables the representation of morphological patterns for each entry (a MWE, a word or an affix). The connection of a lexical entry to an ontological entity is marked mainly by the `denotes` property or is mediated by the `LexicalSense` or the `LexicalConcept` properties, as this is represented in Figure 1, which displays the core module of the model.

As stated in Section 1, OntoLex-Lemon builds on and extends the *lemon* model. A major difference is that OntoLex-Lemon includes an explicit way to encode conceptual hierarchies, using the SKOS standard.¹⁶ As can be seen

¹⁵See (McCrae et al., 2012), (Cimiano et al., 2016) and also https://www.w3.org/community/ontolex/wiki/Final_Model_Specification.

¹⁶SKOS stands for “Simple Knowledge Organization Sys-

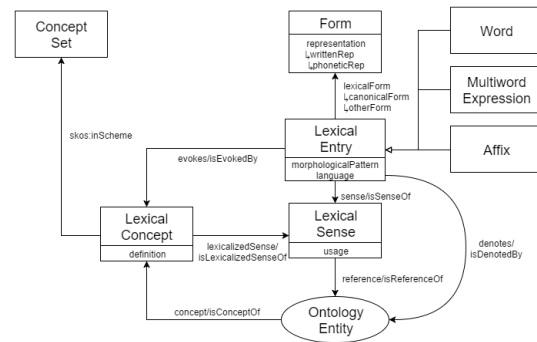


Figure 1: The core module of OntoLex-Lemon: Ontology Lexicon Interface. Graphic taken from <https://www.w3.org/2016/05/ontolex/>.

in Figure 1, lexical entries can be linked, via the `ontolex:evokes` property, to such SKOS concepts, which can represent WordNet synsets. This structure is paralleling the relation between lexical entries and ontological resources, which is implemented either directly by the `ontolex:reference` property or mediated by the instances of the `ontolex:LexicalSense` class.¹⁷ The `ontolex:LexicalConcept` class seems to be best appropriated to model the “sets of cognitive synonyms (synsets)”¹⁸ that Princeton WordNet (PWN) describes, while the `ontolex:LexicalSense` class is meant to represent the bridge between lexical entries and ontological entities (which do not necessarily have semantic relations between them).

4 Mapping OdeNet to OntoLex-Lemon

A main issue with the original partly crowd-sourced data for OdeNet was that additional textual information or special characters were added

tem”. SKOS provides “a model for expressing the basic structure and content of concept schemes such as thesauri, classification schemes, subject heading lists, taxonomies, folksonomies, and other similar types of controlled vocabulary” (<https://www.w3.org/TR/skos-primer/>)

¹⁷Quoting from Section 3.6 “Lexical Concept” <https://www.w3.org/2016/05/ontolex/>: “We [...] capture the fact that a certain lexical entry can be used to denote a certain ontological predicate. We capture this by saying that the lexical entry denotes the class or ontology element in question. However, sometimes we would like to express the fact that a certain lexical entry evokes a certain mental concept rather than that it refers to a class with a formal interpretation in some model. Thus, in *lemon* we introduce the class `Lexical Concept` that represents a mental abstraction, concept or unit of thought that can be lexicalized by a given collection of senses. A lexical concept is thus a subclass of `skos:Concept`.”

¹⁸Quoted from <https://wordnet.princeton.edu/>.

by the crowd to the headwords. In order to clean the data, we wrote a Python script, which not only is filtering out noisy data, but also mapping certain GWN codes (like part of speech (PoS)) into the vocabularies used in OntoLex-Lemon, like for example the LexInfo vocabulary for PoS and semantic relations.¹⁹

As for now, we have in the OntoLex-Lemon encoding of OdeNet 120,012 lexical entries, the same number of lexical senses and 36,192 synsets, which are encoded as instances of the class `ontolex:LexicalConcept` and included in a SKOS-based conceptual hierarchy, supporting also the description of lexical semantic relations between synsets, like synonymy, hyponymy etc. It is interesting to notice that 44,506 entries contain a blank and can therefore be considered as Multi Word Expressions. And if we add to this figure all the 14,080 compound entries²⁰ we note that approximately half of the lexical entries in the OntoLex-Lemon representation can be considered as MWEs.

The following listings give some details on the OntoLex-Lemon encoding of the first entry in OdeNet, which is “Kernspaltung” (*nuclear fission*).

Listing 1: The lexical entry for *Kernspaltung*

```
:entry_w1
rdf:type ontolex:MultiWordExpression ;
decomp:constituent :Kern_comp ;
rdf:_1 :Kern_comp ;
decomp:subterm :entry_w3542 ;
decomp:constituent :spaltung_comp ;
rdf:_2 :spaltung_comp ;
decomp:subterm :entry_w23527 ;
lexinfo:hypernym :synset_odenet-5437-n ;
wn:partOfSpeech wn:noun ;
ontolex:canonicalForm :form_w1 ;
ontolex:sense :sense_w1_1-n ;
ontolex:evokes :synset_odenet-1-n ;
.
```

In Listing 1 we display the full OntoLex-Lemon entry. One aspect that can be immediately noted by the reader, is the possibility to represent the components of the compound word, which is encoded as being an instance of

¹⁹See <https://www.lexinfo.net/ontology/2.0/lexinfo> and also (Cimiano et al., 2011).

²⁰This figure was computed merely by comparison with the list of split nominal compounds offered by the GermaNet project on its web page: http://www.sfs.uni-tuebingen.de/GermaNet/documents/compounds/split_compounds_from_GermaNet13.0.txt. We expect to have a larger number of compounds by applying a decomposition algorithm, not only to nominal entries.

the class `ontolex:MultiWordExpression` (which in OntoLex-Lemon marks any type of entries that can be segmented, thus including compounds). This possibility is demonstrating one of the added-value of linking synsets to the (complex) representation of lexical entries, as we can state (see below) semantic relations between synsets associated to the components of a compound word and its synsets.

Listing 2 below is displaying the form information associated to the `w1` entry in Listing 1.

Listing 2: The `ontolex:Form` *Kernspaltung*

```
:form_w1
rdf:type ontolex:Form ;
ontolex:writtenRep "Kernspaltung"@de ;
.
```

Listing 3 is showing the conversion of the original OdeNet sense information into an instance of the `ontolex:LexicalSense` class.

Listing 3: The `LexicalSense` associated to the entry for *Kernspaltung*

```
:sense_w1_1-n
rdf:type ontolex:LexicalSense ;
ontolex:isLexicalizedSenseOf
:synset_odenet-1-n ;
ontolex:isSenseOf :entry_w1 ;
ontolex:reference
https://www.wikidata.org/wiki/Q11429 ;
.
```

In this code we can see how a sense can be linked to a synset, via the property `ontolex:isLexicalizedSenseOf`, while the entry itself can be linked to the synset via the property `ontolex:evokes`, as this is displayed in Listing 1. The sense itself is also linking (`ontolex:reference`) to an ontological entity, here in the form of a Wikidata entry.

Listing 4 displays the representation of the synset associated to both the `w1` lexical entry and the `w1_1-n` sense. There we can also see that this lexical concept (synset) is also “evoked” by other entries/senses. For example by the entries for “Kernfission” or “Atomspaltung”, which are synonyms of “Kernspaltung”. The `lexinfo:hypernym` property is providing the information on the semantic relation this synset has to another synset.

Listing 4: The `LexicalConcept` (synset) associated to the entry for *Kernspaltung*

```
:synset_odenet-1-n
rdf:type ontolex:LexicalConcept ;
skos:inScheme :ODenet ;
skos:definition "a nuclear reaction
```

```

in which a massive nucleus splits
into smaller nuclei with the
simultaneous release of energy” ;
wn:ili ili:il107577 ;
ontolex:isEvokedBy :entry_w1 ;
ontolex:isEvokedBy :entry_w2 ;
ontolex:isEvokedBy :entry_w3 ;
ontolex:isEvokedBy :entry_w4 ;
ontolex:lexicalizedSense :sense_w1_1-n ;
ontolex:lexicalizedSense :sense_w2_1-n ;
ontolex:lexicalizedSense :sense_w3_1-n ;
ontolex:lexicalizedSense :sense_w4_1-n ;
lexinfo:hypernym :synset_odenet-5437-n ;
.

```

Finally, in Listing 5 we display the “entries” for the components of the compound word “Kernspaltung”. Those components are pointing to the lexical entries they are related to (the entry `:entry_w23527` is for example the one corresponding to the noun “Spaltung” (*split, fission, separation, cleavage*, etc.), which has again its own senses and associated synsets. We can here disambiguate the meaning of “Spaltung” as used in the compound, as being the one of “fission”. And the whole compound can then be considered as an hyponym of the synset for “fission”.

Listing 5: The two components of the entry *Kernspaltung*

```

:Kern_comp
  rdf:type decomp:Component ;
  decomp:correspondsTo :entry_w3542 ;
.
:spaltung_comp
  rdf:type decomp:Component ;
  decomp:correspondsTo :entry_w23527 ;
.

```

In Listing 1 above, we can see the information on the sequence those components have in this entry. For sure, those component “entries” can be re-used separately for other compound, like for example for “Atomspaltung”. So that we can collect all the corresponding meanings of a word, also when they are used in compounds, also in dependency of their position in the compounds. Details on the decomposition module of OntoLex-Lemon are shown in Figure 2.

In this section we described the current state of the OntoLex-Lemon representation of the data we can find in the OdeNet resource. But we also touched the possible use of OntoLex-Lemon for bridging WordNet-like resources and full lexical descriptions, concentrating in the above section on the topic of German compound nouns. In the next section we present the morphological resource we mapped onto OntoLex-Lemon in order to be able

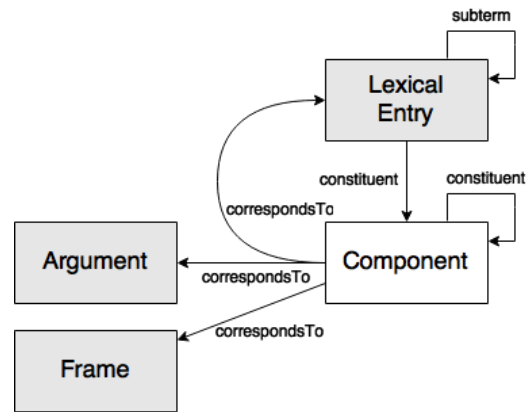


Figure 2: The Decomposition module of OntoLex-Lemon. Graphic taken from <https://www.w3.org/2016/05/ontolex/>.

to link OdeNet elements to a full morphological description.

5 MMORPH

As mentioned in Section 1 we work with an updated German version of MMORPH (Petit-pierre and Russell, 1995), which covers also English, Spanish, French and Italian morphology. Our German version of MMORPH contains over 2,630,000 full-forms, and has specifically improved the coverage of compounds compared to the original German version of MMORPH. MMORPH presents its data in a well structured fashion, as the (simplified) example for the noun “Kernspaltung” (*nuclear fission*) below demonstrates:

Listing 6: The MMORPH entry for *Kernspaltung*

```

" kernspaltung" =
  " kernspaltung" Noun[ gender=fem
    number=singular case=nom|gen|dat|acc ]
" kernspaltungen" =
  " kernspaltung" Noun[ gender=fem
    number=plural case=nom|gen|dat|acc ]

```

We wrote a script in order to transform the MMORPH data into OntoLex-Lemon, in its turtle syntax serialization.²¹ We made use for this of the Python `rdflib` module,²² which supports the generation of RDF-graphs in `rdf:xml`, `json-ld` or `turtle` formats. As for nouns, out of 349,874 original full forms 67,717 instances of the `ontolex:LexicalEntry` class have been

²¹More on the turtle syntax: <https://www.w3.org/TR/turtle/>.

²²See <https://github.com/RDFLib/rdflib> for more details.

generated, and 224,439 instances of the class `ontolex:Form`. 67,717 forms are referred to as a `ontolex:canonicalForm` (one for each lexical entry). We have 27,387 adjectives as instances of the `ontolex:LexicalEntry` class, and 474,459 instances of the class `ontolex:Form`. This figure shows the high number of morphological variants for adjectives in German (447,072 forms are marked with the property `ontolex:otherForm`).

6 Linking the OdeNet Resource to the MMORPH Resource

We see the use of OntoLex-Lemon for representing WordNets as a chance to not only port information from one format to another (with the possibility to publish WordNets in the Linguistic Linked Open Data cloud), but also as an opportunity to extend the coverage of WordNet descriptions to more complex lexical phenomena, beyond lemma and PoS considerations. One case we have been studying concerns the morphological specification of MWEs included in OdeNet.

As we could see, there are more than a significant number of MWEs in OdeNet, being compounds or “phrasal entries”, like for example “Rotkohl” (*purple cabbage* or *red cabbage*), “Rotkraut” (*red kraut* or *red cabbage*), “rote Bete” (*beetroot*), or “geistiges Eigentum” (*intellectual property*). A note on those examples: While “Rotkraut” and “Rotkohl” are essentially pointing to the same vegetable,²³ the word “Rotkraut” is typically used only in its singular form.²⁴ The same remark for the MWE “geistiges Eigentum”.²⁵

There is no way in the original OdeNet (or in general in PWN or other WordNets) to explicitly formulate this restriction, that an entry can be used

²³See for this the OpenThesaurus: <https://www.openthesaurus.de/synonyme/Rotkohl> or Duden: <https://www.duden.de/rechtschreibung/Rotkohl>.

²⁴See for this Wiktionary: <https://de.wiktionary.org/wiki/Rotkraut> or Duden: <https://www.duden.de/rechtschreibung/Rotkraut> where only the singular forms are listed. Wiktionary indicates a plural for “Rotkohl” (<https://de.wiktionary.org/wiki/Rotkohl>).

²⁵It is interesting to note that neither Duden nor CanooNet have an entry for such MWEs like “rote Bete” or “geistiges Eigentum”, but Wiktionary and OdeNet include such MWEs as entries. We guess that a lexicography view is dedicated to include only words, also resulting from word formation processes, as entries, while the other dictionary tradition is more closely related to the description of meanings.

only in singular. It is possible in PWN though to get the information that a concept is only lexicalized by a plural form, by just querying for a plural form, like for example “peoples”. If this plural form is not reduced exclusively to its lemma, then a synset for it will be returned, together with the synsets for the singular form, as can be seen in the following listing, where the plural form is highlighted²⁶:

Listing 7: The Synsets for “people” vs. “peoples”

```

people.n.01 ((plural) any group of human
  beings ... collectively)
citizenry.n.01 (the body of citizens of
  a state or country)
people.n.03 (members of a family line)
multitude.n.03
(the common people generally)
peoples.n.01 (the human beings of a
  particular nation or community or
  ethnic group)

```

But it is to be noticed, that when the user is querying for “people”, the synset for the plural form “peoples” will not be displayed.

The example of the OntoLex-Lemon representation of the German adjective “rot” (*red*) displayed in Listing 8 is introduced in order to give an idea of the complexity of the inflectional variants for a German adjective, whereas we do not include the form variants that are conditioned by the preceding use of a definite or an indefinite determiner.

Listing 8: The MMORPH entry for *rot* (red) in OntoLex-Lemon

```

:lex_rot a ontolex:LexicalEntry ;
lexinfo:partOfSpeech lexinfo:adjective ;
ontolex:canonicalForm :form_rot ;
ontolex:otherForm
:form_rot_comp_dat_neutrum-masc_singular ,
:form_rot_comp_gen-dat-acc_singular ,
:form_rot_comp_gen_plural ,
:form_rot_comp_nom-acc_neutrum_singular ,
:form_rot_comp_nom-acc_plural ,
:form_rot_comp_nom-acc_singular ,
:form_rot_comp_nom-gen-dat_masc-fem_singular ,
:form_rot_comp_plural ,
:form_rot_pos_dat_neutrum-masc_singular ,
:form_rot_pos_gen-dat-acc_singular ,
:form_rot_pos_gen_plural ,
:form_rot_pos_nom-acc_neutrum_singular ,
:form_rot_pos_nom-acc_plural ,
:form_rot_pos_nom-acc_singular ,
:form_rot_pos_nom-gen-dat_masc-fem_singular ,
:form_rot_pos_plural ,
:form_rot_sup_dat_neutrum-masc_singular ,
:form_rot_sup_gen-dat-acc_singular ,
:form_rot_sup_gen_plural ,
:form_rot_sup_nom-acc_neutrum_singular ,
:form_rot_sup_nom-acc_plural ,
:form_rot_sup_nom-acc_singular ,
:form_rot_sup_nom-gen-dat_masc-fem_singular ,
:form_rot_sup_plural .

```

Listing 9 displays the morphological variants

²⁶This listing was generated from the user interface of Princeton WordNet: <http://wordnetweb.princeton.edu/perl/webwn>

of the noun “Bete” (*beet*)²⁷, as represented in OntoLex-Lemon.

Listing 9: The MMORPH entry for *Bete* (beet) and its 2 possible ontolex:Forms in the OntoLex-Lemon representation

```
:lex_bete a ontolex:LexicalEntry ;
  lexinfo:gender lexinfo:feminine ;
  lexinfo:partOfSpeech lexinfo:noun ;
  ontolex:canonicalForm :form_bete ;
  ontolex:otherForm :form_bete_plural .

:form_bete a ontolex:Form ;
  lexinfo:case lexinfo:nominative ;
  lexinfo:number lexinfo:singular ;
  ontolex:writtenRep "Bete"@de .

:form_bete_plural a ontolex:Form ;
  lexinfo:number lexinfo:plural ;
  ontolex:writtenRep "Beten"@de .
```

But even if we can limit the number of forms for the noun “Bete”, we have to combine those with the possible forms of “rot”, and also consider the possible use of a preceding indefinite or definite determiner. This gives us 32 forms to be considered, compared to the maximum of 8 different forms if we deal with a nominal compound like “Rotkohl”. And in fact, the OntoLex-Lemon linking mechanisms allow us to precise only the “positive” adjectival forms, as “rot” can not appear in this MWE as a comparative or superlative.

We have a similar situation with the entry “geistiges Eigentum” (*intellectual property*)²⁸ in OdeNet. But there is another restriction, following which this concept of intellectual property can be used only in the singular form. We select from the list of 24 possible forms of the adjective only the “positive” ones (see Listing 10).

Listing 10: The MMORPH positive forms for *geistig* (intellectual/spiritual) in the OntoLex-Lemon representation

```
:lex_geistig a ontolex:LexicalEntry ;
  lexinfo:partOfSpeech lexinfo:adjective ;
  ontolex:canonicalForm :form_geistig ;
  ontolex:otherForm
    :form_geistig_pos_dat_neutrum-masc_singular ,
    :form_geistig_pos_gen-dat-acc_singular ,
    :form_geistig_pos_gen_plural ,
    :form_geistig_pos_nom-acc_neutrum_singular ,
    :form_geistig_pos_nom-acc_plural ,
    :form_geistig_pos_nom-acc_singular ,
    :form_geistig_pos_nom-gen-dat-masc-fem_singular ,
    :form_geistig_pos_plural .
```

The OntoLex-Lemon representation of the MMORPH entry for the noun “Eigentum”, with

²⁷The number of forms can be reduced as all forms in singular have the same ending, the same for the plural, so that we do not need to list all different grammatical cases.

²⁸We note that “intellectual property” is also a MWE entry in PWN WordNet.

the links to the associated forms, which are not displayed here, is shown in Listing 11

Listing 11: The MMORPH noun *Eigentum* (property) with the corresponding form variants

```
:lex_eigentum
  a ontolex:LexicalEntry ;
  lexinfo:partOfSpeech lexinfo:noun ;
  ontolex:canonicalForm :form_eigentum ;
  ontolex:otherForm
    :form_eigentum_dat_plural ;
  ontolex:otherForm
    :form_eigentum_gen_singular ;
  ontolex:otherForm
    :form_eigentum_nom_gen_acc_plural ;
  .
```

One possibility would be to link the OdeNet entry “geistiges Eigentum” to both the relevant forms displayed just above in the Listings 10 and 11, with an additional information on the word ordering and that only the singular forms can be selected. This can be ensured by the use of the `ontolex:usage` property. This solution has the advantage that we do not have to introduce the phrasal MWE in the MMORPH representation corresponding to the OdeNet “geistiges Eigentum”. But at the price of introducing some rules, like the ordering of the words, an agreement rule or the specific restriction that an OdeNet entry has only singular forms for its lexicalization.

The other possibility is to introduce an entry for the OdeNet MWE and the corresponding forms, as this is shown in Listing 12

Listing 12: The OdeNet entry *geistiges Eigentum* (intellectual property) pointing to all possible (singular) forms

```
:lex_geistiges_eigentum
  a ontolex:MultiWordExpression ;
  lexinfo:partOfSpeech lexinfo:noun ;
  <http://www.w3.org/ns/lemon/decomp#subterm>
    :lex_eigentum ;
  <http://www.w3.org/ns/lemon/decomp#subterm>
    :lex_geistig ;
  rdf:_1 :lex_geistig ;
  rdf:_2 :lex_eigentum ;
  ontolex:canonicalForm
    :form_geistiges_eigentum ;
  ontolex:otherForm
    :form_geistiges_eigentum_dat_singular ;
  ontolex:otherForm
    :form_geistiges_eigentum_dat_singular_def_indef ;
  ontolex:otherForm
    :form_geistiges_eigentum_nom_acc_singular_def ;
  ontolex:otherForm
    :form_gesitiges_eigentum_gen_singular ;
  ontolex:usage lexinfo:singular ;
  .
```

In this example we included the decomposition information, including the ordering of the components. We also included the line `ontolex:usage lexinfo:singular`, which can be considered as redundant as we already selected as `ontolex:otherForm` all

the singular forms, discarding thus the (possible) plural forms.

As a purely morphological information source, MMORPH does not have any sense or synset associated with its entries. Linking them to the OdeNet resources is adding thus a conceptual view to the MMORPH data. Additionally one can add reference information by querying DBpedia²⁹ or Wikidata³⁰ This can be done very easily by just adding the line `ontolex:denotes` with the corresponding URL pointing to an ontological reference.

7 Conclusion

We described our current work consisting in porting a recently developed German WordNet compliant lexical resource, OdeNet, to OntoLex-Lemon, in order to support its publication in the Linguistic Linked Open Data cloud. While processing those data, we noticed that OntoLex-Lemon can be used for bridging the WordNet type of lexical resources to a full description of lexical entries, leading to an extension of the coverage of WordNets beyond the consideration of lemmas and PoS information. In order to test our intuition, we ported an updated version of the German MMORPH morphological analyzer to OntoLex-Lemon and we established links between the two new OntoLex-Lemon data sets. We documented our interlinking work with the example of the full morphological representation of components of German compounds and MWEs used in OdeNet, also being able to express usage restrictions.

Acknowledgments

Contributions by Thierry Declerck have been supported in part by the H2020 project “ELEXIS” with Grant Agreement number 731015 and by the H2020 project “Prêt-à-LLOD” with Grant Agreement number 825182.

²⁹See <https://wiki.dbpedia.org/>. “DBpedia ... is a project aiming to extract structured content from the information created in the Wikipedia project”, quoted from <https://en.wikipedia.org/wiki/DBpedia>.

³⁰“Wikidata is a free and open knowledge base that can be read and edited by both humans and machines. Wikidata acts as central storage for the structured data of its Wikimedia sister projects including Wikipedia, Wikivoyage, Wiktionary, Wikisource, and others.”, quoted from https://www.wikidata.org/wiki/Wikidata:Main_Page.

References

- Francis Bond and Ryan Foster. 2013. [Linking and extending an open multilingual WordNet](#). In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 1352–1362, Sofia.
- Francis Bond and Kyonghee Paik. 2012. A survey of WordNets and their licenses. *Small*, 8(4):5.
- Philipp Cimiano, Paul Buitelaar, John McCrae, and Michael Sintek. 2011. Lexinfo: A declarative model for the lexicon-ontology interface. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, 9(1):29–51.
- Philipp Cimiano, John P. McCrae, and Paul Buitelaar. 2016. Lexicon Model for Ontologies: Community Report.
- Judith Eckle-Kohler, John Philip McCrae, and Christian Chiarcos. 2015. [lemonUby - a large, interlinked, syntactically-rich lexical resource for ontologies](#). *Semantic Web Journal*, 6(4):371–378.
- Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.
- Birgit Hamp, Helmut Feldweg, et al. 1997. GermaNet- a lexical-semantic net for German. In *Proceedings of ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, pages 9–15.
- Matthias Hüning and Barbara Schlücker. 2015. *Word-Formation. An International Handbook of the Languages of Europe*, chapter Multi-word expressions. De Gruyter Mouton, Berlin, Boston.
- John McCrae, Guadalupe Aguado-de Cea, Paul Buitelaar, Philipp Cimiano, Thierry Declerck, Asuncion Gomez-Perez, Jorge Garcia, Laura Hollink, Elena Montiel-Ponsoda, Dennis Spohr, and Tobias Wunner. 2012. Interchanging Lexical Resources on the Semantic Web. *Language Resources and Evaluation*, 46(4):701–719.
- John P. McCrae, Paul Buitelaar, and Philipp Cimiano. 2017. The OntoLex-Lemon Model: Development and Applications. In *Proceedings of eLex 2017*, pages 587–597. INT, Trojina and Lexical Computing, Lexical Computing CZ s.r.o.
- John P. McCrae, Christiane Fellbaum, and Philipp Cimiano. 2014. [Publishing and linking WordNet using lemon and RDF](#). In *Proceedings of the 3rd Workshop on Linked Data in Linguistics*.
- Dominique Petitpierre and Graham. Russell. 1995. [MMORPH: The Multext morphology program](#). Multext deliverable 2.3.1, ISSCO, University of Geneva.