Towards an RDF Representation of the Infrastructure consisting in using Wordnets as a conceptual Interlingua between multilingual Sign Language Datasets

Thierry Declerck
DFKI GmbH
Saarland Informatics Campus
Stuhlsatzenhausweg, 3
66123 Saarbrücken
declerck@dfki.de

Thomas Troelsgård
University College
Copenhagen (KP)
Humletorvet 3
1799 Copenhagen V
ttro@kp.dk

Sussi Olsen
Centre for Language Technology,
NorS, University of Copenhagen
Emil Holms Kanal 2
2300 Copenhagen S
saolsen@hum.ku.dk

Abstract

We present ongoing work dealing with a Linked Data compliant representation of infrastructures using wordnets for connecting multilingual Sign Language data sets. We build for this on already existing RDF and OntoLex representations of Open Multilingual Wordnet (OMW) data sets and work done by the European EASIER research project on the use of the CSV files of OMW for linking glosses and basic semantic information associated with Sign Language data sets in two languages: German and Greek. In this context, we started the transformation into RDF of a Danish data set, which links Danish Sign Language data and the wordnet for Danish, DanNet. The final objective of our work is to include Sign Language data sets (and their conceptual cross-linking via wordnets) in the Linguistic Linked Open Data cloud.

1 Introduction

A final goal of our work is to represent and publish Sign Language (SL) data sets in the Linguistic Linked Data (LLOD) cloud, which is a subset of the Linked Data (LD) cloud.[1] We can observe that SL data are not represented in the data sets currently included in the LLOD cloud. And looking at the "Overview of Datasets for the Sign Languages of Europe" published by the "EASIER" European project (Kopf et al., 2022)[2] we do not see any mention of a data set being available in a Linked Data compliant format.

This shortcoming is a problematic issue, as an important type of natural language is missing from the LLOD, while the motivation behind the creation of the LLOD is that it can ease the linking of all types of natural language resources.[3]

The prerequisite for publishing linguistic data in the LLOD cloud is to have it formally represented within the Resource Description Framework (RDF).[4] And as an RDF-based de facto standard for representing lexical information, the OntoLex-Lemon specifications,[5] already exists, we investigate as a first step the re-use of this model in order to accommodate the description of Sign Language data sets. But as we can see in Figure 1, the class `ontolex:Form` covers only the representation of written languages (with the addition of the associated pronunciation information), so that there is a need to think about possible adaptations or extensions of OntoLex-Lemon.

At the same time, the OntoLex-Lemon model supports the representation of WordNet data, which are typically encoded with the SKOS[6] vocabulary, where the synsets are represented as instances of the `ontolex:LexicalConcept` subclass of the `skos:Concept` class.[7] This feature is offering us a good starting point for transforming into RDF (and OntoLex-Lemon) recent work by the EASIER project dealing with the use of shared IDs of the Open Multilingual Wordnet (OMW)[8] infrastructure for interlinking SL data sets for two languages: German and Greek, as described in (Bigeard et al., 2022).[9]

[1] Those clouds can be accessed respectively at http://linguistic-lod.org/llod-cloud and https://lod-cloud.net/

[2] Available as a public deliverable at https://www.project-easier.eu/deliverables/

[3] See (Chiarcos et al., 2012) for a first description of the motivations leading to the creation of the LLOD, and (Cimiano et al., 2020) for a more recent and much more detailed description of all aspects of the LLOD infrastructure

[4] See https://www.w3.org/TR/rdf11-primer/ for an introduction to RDF

[5] See https://www.w3.org/2016/05/ontolex/ and (McCrae et al., 2017)

[6] SKOS stands for "Simple Knowledge Organization System". see https://www.w3.org/TR/skos-primer/ for more details

[7] See for example (Declerck, 2019)

[8] See (Bond and Foster, 2013) and (Bond et al., 2016) for more details on the Open Multilingual Wordnet and the interlinking between OMW data sets

[9] The EASIER project is publishing the related data at https://www.fdr.uni-hamburg.de/record/10169#.Y1Ufs-RBzmF

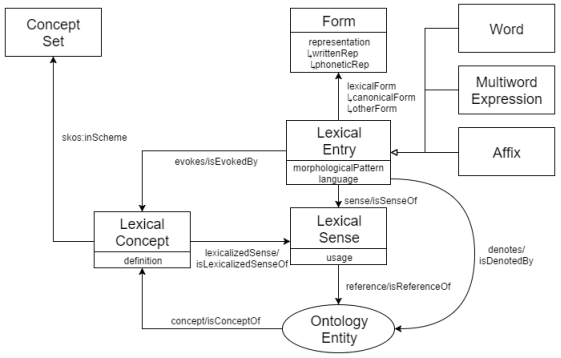

Figure 1: The core module of OntoLex-Lemon, taken from https://www.w3.org/2016/05/ontolex/

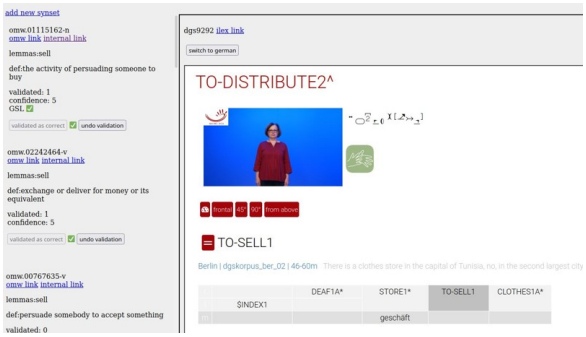

Figure 2: A screenshot showing how German and Greek Sign Language data are interlinked via a shared OMW index, as proposed by the EASIER project. Taken from https://www.fdr.uni-hamburg.de/record/10169#.Y01WXExBzmE, screenshot4.jpg

Figure 2,[10] displays the starting point of a video representing a sign, its related glosses (written using only capital letters, in order to distinguish them from keywords or potential lexical entries),[11] its phonetic transcription (in HamNoSys, (Hanke, 2004)), as well as its link to the OMW version of the Princeton WordNet (omw.01115162-n, with its lemma and definition).

This link to OMW is done for the German sign indexed with "dgs9292",[12] which points to the subtype (or the "sub-gloss") "TO-SELL1" of "TO-DISTRIBUTE2^". Those English glosses are in fact translations of the German glosses "VERKAUFEN1" and "VERTEILEN2^", which are used for accessing the Princeton WordNet in OMW (as the GermaNet resource (Kunze and Lemnitzer, 2002) used in EASIER is not included in OMW). It can also be seen in Figure 2 that a "GSL" box is being positively checked. "GSL" stands for "Greek Sign Language", and the positively checked abbreviation in the screenshot means that there is a corresponding synset in the Greek Wordnet available in OMW. This way, a DGS sign can be linked to a GSL sign, based on a shared OMW ID, which is much more accurate than linking only via translation of glosses.

The links between the one OMW ID and the two signs/videos IDs are available in Excel files.[13] The corresponding CSV lines are displayed in Figure 3 and Figure 4, where we can see that one OMW ID (omw.00377364-n, with the English lemma "explosion", translated to German "*Explosion*", and with the Greek lemma "*έκρηξη*") is associated with both the German and the Greek SL resources, thus establishing a conceptual link between those.

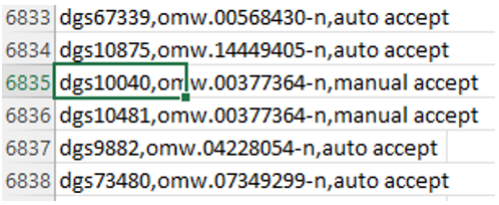

Figure 3: The CSV representation of the linking of OMW and a German sign, taken from https://www.fdr.uni-hamburg.de/record/10169#.Y01WXExBzmE

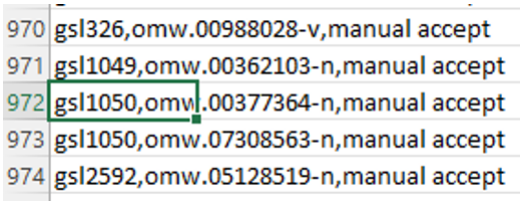

Figure 4: The CSV representation of the linking of OMW and a Greek sign, https://www.fdr.uni-hamburg.de/record/10169#.Y01WXExBzmE

Those elements: videos, glosses, phonetic transcriptions, links to OMW, are the elements we are encoding in a unified and harmonized Linked Data compliant format.

[10] As stated in the web page, from which this screenshot is taken, the online interface is not yet live. But the displayed screenshot represents clearly how the linking of a German sign and an OMW ID is (will be) represented on the web.

[11] On the specificity of glosses used for naming (or labelling) SL data in corpora, see (Ormel et al., 2010)

[12] DGS stands for "Deutsche Gebärdensprache", *German Sign Language*

[13] Also made available at https://www.fdr.uni-hamburg.de/record/10169#.Y01WXExBzmE

2 Linked Data compliant Encoding of the Infrastructure using shared OMW IDs

As stated in the introduction, we need to transform into RDF the different types of data used for representing signs for their future publication in the LLOD cloud. We also make use of RDF(S) and OWL representation languages, as those are constitutive parts of the OntoLex-Lemon specifications and of the building of ontologies.[14]

For the RDF representation of videos included in our data set, we just introduce a class and have all videos encoded as instances of this class.

Listing 1 displays the RDF-based encoding of a video containing a German sign.[15] A partial view of the original web page is displayed in Figure 5.

Figure 5: The German sign associated with the gloss SCHUTZ1A^ in the DGS Corpus https://www.sign-lang.uni-hamburg.de/meinedgs/types/type13990_de.html

In the RDF representation of the sign, it can be seen that the video/sign is linked to two glosses, as this sign has more than one gloss related to it.

Listing 1: The RDF-based encoding of a video containing a sign

```
<http://example.org/dgs#
        SignVideos_40085921.mp4>
   rdf:type sl:SignVideos ;
   dgs:hasGLOSS dgs:GLOSS_13990 ;
   dgs:hasGLOSS dgs:GLOSS_13990-2966 ;
   sl:hasVideoAdresss
         "https://www.sign-lang.
              uni-hamburg.de/korpusdict/
              clips/4008592_1.mp4"^^rdf:HTML ;
   rdfs:label "\"Videos representing
                 a sign\"" ;
.
```

Listing 2 displays the corresponding glosses (as instances of a specific class).

Listing 2: The RDF-based encoding of glosses

```
dgs:GLOSS_13990
   rdf:type dgs:GLOSS ;
   rdfs:label "\"PROTECTION1A^\""@en ;
   rdfs:label "\"SCHUTZ1A^\""@de ;
.
dgs:GLOSS_13990-2966
   rdf:type dgs:GLOSS ;
   rdfs:label "\"PROTECTION1A\""@en ;
   rdfs:label "\"SCHUTZ1A\""@de ;
.
```

The subclass/subtype relation between the glosses displayed in Listing 2 is encoded in a specific class, called "Type", as can be seen in Listing 3, which displays the subclass hierarchy between glosses (here class and subclass instances are linking to the same video), and linking the instance of the subclass to an OMW element (as instance of the class `ontolex:LexicalConcept`), establishing thus the link to the WordNet world, and to the corresponding video(s), as we can collect more than one video representing a sign.

Listing 3: Subclass hierachy of glosses linking the subgloss to OWM and to videos

```
dgs:Type_13990
   rdf:type dgs:Type ;
   dgs:hasGLOSS dgs:GLOSS_13990 ;
   dgs:hasSubType
      dgs:Subtype_13990-2966 ;
   dgs:hasVideo
      <http://example.org/dgs#
      SignVideos_40085921.mp4> ;
   dgs:hasVideo
      <http://example.org/dgs#
       SignVideos_dgs-688.mp4> ;
   rdfs:label "\"Schutz\""@de ;
   rdfs:label "\"protection\""@en ;
.

dgs:Subtype_13990-2966
   rdf:type dgs:Subtype ;
   dgs:hasGLOSS dgs:GLOSS_13990-2966 ;
   dgs:hasOMW-Link wnid:omw-00817680-n ;
   dgs:hasVideo
     <http://example.org/dgs
      #SignVideos_40085921.mp4> ;
   dgs:hasVideo
     <http://example.org/dgs
      #SignVideos_dgs-688.mp4> ;
   rdfs:label "\"Schutz\""@de ;
   rdfs:label "\"protection\""@en ;
.
```

[14] RDF(S) stands for "RDF-Schema", see https://www.w3.org/TR/rdf-schema/ for more details. OWL stands for "Web Ontology Language", see https://www.w3.org/TR/2012/REC-owl2-primer-20121211/ for more details.

[15] The original sign and all the related information are accessible at https://www.sign-lang.uni-hamburg.de/meinedgs/types/type13990_de.html (for the English translation of the page: https://www.sign-lang.uni-hamburg.de/meinedgs/types/type13990_en.html)

The OMW synset linked to in Figure 3 has the internal organisation displayed in Figure 6. Here we didn't include links to glosses or videos, as the relations to OMW described in listing 3 are inverse.

```
wnid:omw-00817680-n
  rdf:type ontolex:LexicalConcept ;
  sl:hasWnLemma "\"protection\""@en ;
  sl:hasWnLemma "\"προστασία\""@el ;
  rdfs:label "\"Schutz\""@de ;
  rdfs:label "\"protection\""@en ;
  rdfs:label "\"προστασία\""@el ;
  skos:definition "\"παρεχόμενη φροντίδα
    σε κάποιον ώστε να προφυλάσσεται από
    υπαρκτούς ή διάφορους πιθανούς
    κινδύνους\""@el ;
  skos:inScheme sl:ConceptSet_OMW-DGS ;
  wn:definition "\"a covering that is
    intend to protect from damage or
    injury\""@en ;
.
```

Figure 6: The ID omw-00817680-n of OMW

Finally, the representation of the form(s) of the sign is performed for the time being as instances of `ontolex:Form` (mediated, also for the time being, by an underspecified instance of `ontolex:LexicalEntry`). This representation, displayed in Figure 7, includes the machine-readable transcription of the HamNoSys code, in the so-called SiGML XML format (Neves et al., 2020). It also includes potential keywords or lexical entries.

3 Extending the EASIER Approach with additional Sign Videos per Language

We searched for other Sign Language resources in order to extend the approach described in (Bigeard et al., 2022), thus linking SL data and wordnets, and then transforming those SLs-wordnets combinations into RDF and OntoLex-Lemon.

We found a basic lexicon of 1000 concepts associated with SL data in 4 languages, English, French, German and Greek, an outcome of the past Dicta-Sign project (Matthes et al., 2012), which is available at the University of Hamburg at `https://www.sign-lang.uni-hamburg.de/dicta-sign/portal/concepts/concepts_eng.html`. This resource is directly relevant to our purposes, as the included videos are equipped with glosses and HamNoSys transcriptions, as shown in Figure 8.

In Figure 8, we observe that the gloss and the HamNoSys transcription for the German video are identical with those deployed in the

```
dgs:Form_13990
  rdf:type ontolex:Form ;
  dgs:hasVideo
   <http://example.org/dgs#SignVideos
     _40085921.mp4> ;
  dgs:hasVideo
   <http://example.org/dgs#SignVideos
     _dgs-688.mp4> ;
  sl:hasGloss "\"protection\""@en ;
  sl:hasTranslationInSpokenLanguage
   "\"geschützt, Schutz, schützen\""@de ;
  sl:hasTranslationInSpokenLanguage
   "\"protection\""@en ;
  sl:hasTranslationInSpokenLanguage
   "\"προστασία\""@el ;
  rdfs:label "\"protection\""@en ;
  ontolex:representation
   "\"https://www.sign-lang.
     hamburg.de/galex/glossen/
     g13990.html\"" ;
  ontolex:writtenRep
    "\"hamsymmpar,hamparbegin,hamfist,
    hamthumbacrossmod,hamextfingerdo,
    hamextfingerol,hampalml,hamplus,
    hamfist,hamthumbacrossmod,
    hamextfingeror,hampalmdr,
    hamparend,hamparbegin,
    hamthumbside,hambetween,
    hamindexfinger,hamplus,
    hamfingerbase,hamhandback,
    hamparend,hamtouch,hamchest,
    hammovedo\" {hamnosys-sigml}" ;
.
```

Figure 7: The encoding of the form of a sign

data used by the EASIER project for linking SL data and wordnets, as can be seen at `https://www.sign-lang.uni-hamburg.de/meinedgs/types/type13990_de.html`, and which is also shown in Figure 5.

This concordance of gloss and HamNoSys transcriptions[16] not only allows for the association of two videos representing this German sign to one OWM ID,[17] but it also permits the addition of signs in two additional languages, English and French, extending thus the multilingual coverage of the approach described by (Bigeard et al., 2022). We just need to introduce in our RDF representation new video instances (one per language) and to link them to the same OMW ID.

[16] But we can observe that in the one case the gloss is realised as a noun and in the second case as a verb. Signs are often ambiguous with respect to PoS, and in the future we will link the videos to both the nominal and verbal synsets, if both are available in the corresponding wordnet.

[17] As the page `https://www.sign-lang.uni-hamburg.de/dicta-sign/portal/concepts/cs/cs_688.html` is linking to a more detailed lexical description of the sign, with the same gloss and HamNoSys transcription (see `https://www.sign-lang.uni-hamburg.de/galex/glossen/g13990.html`), with another video for the sign, we can in fact have 3 videos for this German sign associated with one OMW ID.

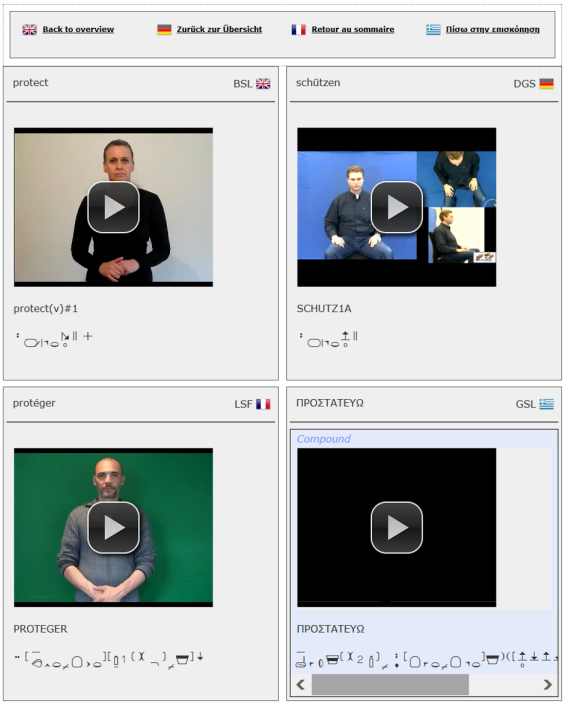

Figure 8: The concept "protect" as realised in 4 different Sign Languages. Taken from https://www.sign-lang.uni-hamburg.de/dicta-sign/portal/concepts/cs/cs_688.html

Thus, the transformation of this additional data into our RDF and OntoLex-Lemon representation means organising those originally disparate and heterogeneous data sources in one harmonised representation format, with OMW as the central component for the interlinking of the different data types and sources.

4 Extending the Approach to Danish WordNet and Sign Language Data

(Troelsgård and Kristoffersen, 2018) discusses approaches for ensuring consistency between (Danish) Sign Language corpus data and the dictionary of Danish signs that is described in (Kristoffersen and Troelsgård, 2010). This approach aims at getting a correspondence between the dictionary lemmas and the corpus lexicon, which consists of types introduced for lemmatising the tokens found in the annotations (glosses added to the signs) of the corpus.

The strategy being to use words and their equivalents (also found in the dictionary) to search for signs in the corpus. In order to extend the list of potential Danish equivalents that could be used for a word-based search of signs in the corpus, (Troelsgård and Kristoffersen, 2018) suggest using the Danish wordnet, DanNet, which is briefly described below and in more details in (Pedersen et al., 2009) and (Pedersen et al., 2018).

DanNet was constructed using the merge approach where the wordnet is built on a monolingual resource, in this case on the corpus-based Danish dictionary Den Danske Ordbog (DDO, (Lorentzen, 2004), see also https://ordnet.dk/ddo), and subsequently linked to PWN. For DanNet this linking was based on the Princeton Core wordnet, a subset containing 5,000 central concepts of English (see the file "core-wordnet.txt" in the folder "LFglosses.standoff-files.zip " under https://wordnetcode.princeton.edu/morpholinks) that were semi-automatically linked to DanNet. These linked elements constitute the part of DanNet available in OMW.[18]

The relations between sign identifiers and lexical elements from both DanNet and other dictionary sources are encoded in a database, out of which we got a CSV export.

In this export, we first have the signs, which are corresponding to entries in the dictionary of signs available at www.tegnsprog.dk.

A second type of data available in the export holds video links and information about the sign form (HamNoSys/SiGML). The HamNoSys notation, though, is rather coarse, as it is generated automatically from the dictionary's phonological descriptions.

A third type of information included in the export is dealing with the senses associated with the signs and their (form) variants.

Lastly, we also have the information from DanNet and PWN. Our work consists thus in porting all those (interlinked) resources to RDF and OntoLex-Lemon, as we did for the German and the Greek data, as presented in Section 2. In the OMW version of DanNet, we find for example the following information"00817680-n lemma beskyttelse", where the lemma corresponds to the OMW English wordnet "00817680-n lemma protection", sharing thus the same ID for the concept of "protection" as the English and Greek wordnets we have in OMW. So that we can add the Danish sign ID (and video), which we got from the database, to our infrastructure. The Danish sign associated with the wordnet

[18] Since 2018, there has been an ongoing effort to link a larger part of DanNet's more than 65,000 synsets to PWN, this time taking departure in the core Danish vocabulary, see (Pedersen et al., 2019).

lemma “beskyttelse” is displayed in Figure 9.

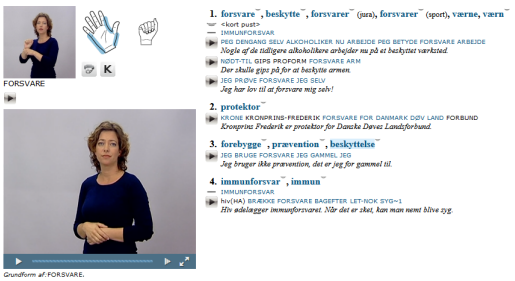

Figure 9: The Danish sign associated with the OMW ID “00817680-n”, corresponding to the (highlighted) lemma “beskyttelse”, here as possible lexical realisation of the Danish gloss “FORSVARE” (*defend*)

It is then straightforward to encode all those types of information on the relation between Danish SL data and DanNet into our RDF-based model. We need only to add an instance for the video displaying the sign, and its associated gloss (with language equivalents), as shown in Figure 10. The language equivalents are included, so that a danish sign can be cross-lingually searched for, using glosses in other languages.

```
dts:GLOSS_dts-1_2162
  rdf:type sl:GLOSS ;
  rdfs:label "\"FORSVARE\""@da ;
  rdfs:label "\"PROTEGER\""@fr ;
  rdfs:label "\"SCHUTZ1A^\""@de ;
  rdfs:label "\"protect(v)#1\""@en ;
  rdfs:label "\"ΠΡΟΣΤΑΤΕΥΩ\""@el ;
.
```

Figure 10: The Danish gloss (with language equivalents) associated with the video with ID dts-1_2162

Then we just have to add an `ontolex:Form` instance for the Danish sign, displayed in Figure 11 and which is linked via its corresponding lexical entry to the corresponding OMW instance, which are shown in Figure 12.

Finally, Figure 13 displays (partially) the current encoding of the OMW ID, showing the central and pivotal role of this ID for interlinking the various types of resources involved in our work.

5 Current Results

Our encoding results in a harmonised representation of data that was originally stored in different formats in different locations. Taking advantage of the work proposed by (Bigeard et al., 2022), (Troelsgård and Kristoffersen, 2018) and others, we can include the links between SL data and

```
dts:Form_1_2162
  rdf:type ontolex:Form ;
  dgs:hasVideo
    <http://example.org/dts#SignVideos_dts-t_2162.mp4> ;
  sl:hasGloss "\"FORSVARE\""@en ;
  sl:hasTranslationInSpokenLanguage "\"forsvare, beskytte,
      forsvarer, værne, værn, beskyttelse\""@da ;
  sl:hasTranslationInSpokenLanguage "\"geschützt, Schutz,
      schützen\""@de ;
  sl:hasTranslationInSpokenLanguage "\"protection\""@en ;
  sl:hasTranslationInSpokenLanguage "\"protection\""@fr ;
  sl:hasTranslationInSpokenLanguage "\"προστασία\""@el ;
  rdfs:label "\"protection, protect\""@en ;
  ontolex:representation "\"https://www.tegnsprog.dk/#
   %7Csoeg%7C'tekst'beskytte%
   7Cresultat%7C4%7Ctrestjerner%7C1%7Ctegn%7C837\"" ;
  ontolex:writtenRep "<sigml><hns_sign gloss='FORSVARE'>
   <hamnosys_manual><hamsymmlr/><hamfist/><hamparbegin/>
   <hamextfingeru/><hampalmd/><hamplus/><hamextfingerr/>
   <hampalmr/><hamparend/><hamparbegin/><hammoveu/>
   <hamthumbside/><hamtouch/><hamplus/><hamnomotion/>
   <hamparend/><hamrepeatfromstart/></hamnosys_manual>
   </hns_sign></sigml>\" {hamnosys-sigml}" ;
.
```

Figure 11: The encoding of OntoLex-Lemon form associated with the sign, where various lexical realisations of the gloss (and of the OMW ID are included, as well as the SiGML code.

```
dts:LexicalEntry_1_2162
  rdf:type ontolex:LexicalEntry ;
  rdfs:label "\"forsvare, beskytte,
    værne, værn, beskyttelse\""@da ;
  ontolex:evokes wnid:omw-00817680-n ;
  ontolex:lexicalForm dts:Form_1_2162 ;
.
```

Figure 12: The encoding of OntoLex-Lemon entry associated with the sign and its ontolex:Form, and which is linking to the corresponding OMW ID

wordnets in a harmonised representation, under the umbrella of RDF and by re-using elements of OntoLex-Lemon. The Open Multilingual Wordnet infrastructure is playing a central role in this work, as the shared OMW IDs across various languages are at the core of the interlinking of the distinct data types and sources. The resulting unified representation supports a dense linking of different types of information. Our model will be made available on Github (https://github.com/Declerck/sl-wn-rdf-ontolex)

6 Future Work

The next steps of our work will consist in automatising the transformation into RDF and aspects of OntoLex-Lemon so that we have all the data in the harmonised representation space. We are also planning to investigate a transformation of ASLNet (Lualdi et al., 2021) into RDF. We continue to extend our work with more data in more languages, starting with Maltese,[19] as a low resourced lan-

[19]For example, a useful dictionary resource for Maltese Sign Language is available at https://mlrs.research.

```
wnid:omw-00817680-n
  rdf:type ontolex:LexicalConcept ;
  sl:hasWnLemma "\"beskyttelse\""@da ;
  sl:hasWnLemma "\"forsorg\""@da ;
  sl:hasWnLemma "\"forsvar\""@da ;
  sl:hasWnLemma "\"protection\""@en ;
  sl:hasWnLemma "\"protection\""@fr ;
  sl:hasWnLemma "\"værn\""@da ;
  sl:hasWnLemma "\"προστασία\""@el ;
  rdfs:label "\"Schutz\""@de ;
  rdfs:label "\"beskyttelse\""@da ;
  rdfs:label "\"protection\""@en ;
  rdfs:label "\"protection\""@fr ;
  rdfs:label "\"προστασία\""@el ;
  skos:definition "\"παρεχόμενη φροντίδα
     σε κάποιον ώστε να προφυλάσσεται από
     υπαρκτούς ή διάφορους πιθανούς
     κινδύνους\""@el ;
  skos:inScheme sl:ConceptSet_OMW-DGS ;
  ontolex:isEvokedBy dgs:LexicalEntry_13990-2966 ;
  ontolex:isEvokedBy dts:LexicalEntry_1_2162 ;
  ontolex:isEvokedBy lsf:LexicalEntry_668-n ;
  wn:definition <http://wordnetweb.princeton.edu/
       perl/webwn?o2=&o0=1&o8=1&o1=1&o7=&o5=&o9=
       &o6=&o3=&o4=&s=protection&i=0&h=0000000#c> ;
.
```

Figure 13: The encoding of the OWM ID, linking to corresponding lexical entries, which again are linking to other elements of our data set, as can be seen in 12 for the Danish case

guage. Finally, we aim at adding other types of visual lexical data, like pictograms, as the links between such data and wordnet have been already investigated, for example in (Schwab et al., 2020).

Acknowledgements

The presented work is pursued in the context of the COST Action NexusLinguarum – European network for Web-centered linguistic data science (CA18209), 731015). We thank Thomas Hanke and Sam Bigeard from the Institute of German Sign Language and Communication of the Deaf (IDGS) at the University of Hamburg for providing links and explanations to data developed in the context of the EASIER project (https://www.project-easier.eu/de/). The work we started dealing with Maltese language is pursued in the context of the LT-BRIDGE project, which has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 952194.

um.edu.mt/resources/lsm. This resource is interesting to us, as it makes use of another transcription system than HamNoSys, the SignWriting system (Sutton, 1991), so that our model will deal also with more than one transcription system.

References

Sam Bigeard, Marc Schulder, Maria Kopf, Thomas Hanke, Kiki Vasilaki, Anna Vacalopoulou, Theodoros Goulas, Athanasia-Lida Dimou, Stavroula-Evita Fotinea, and Eleni Efthimiou. 2022. Introducing Sign Languages to a Multilingual Wordnet: Bootstrapping Corpora and Lexical Resources of Greek Sign Language and German Sign Language. In *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources*, pages 9–15, Marseille, France. European Language Resources Association (ELRA).

Francis Bond and Ryan Foster. 2013. Linking and extending an Open Multilingual Wordnet. In *ACL (1)*, pages 1352–1362. The Association for Computer Linguistics.

Francis Bond, Piek Vossen, John P. McCrae, and Christiane Fellbaum. 2016. CILI: the Collaborative Interlingual Index. In *Proc. of the Global WordNet Conference 2016*.

Christian Chiarcos, Sebastian Hellmann, and Sebastian Nordhoff. 2012. The Open Linguistics Working Group of the Open Knowledge Foundation. In *Linked Data in Linguistics*, pages 153–160. Springer, Heidelberg.

Philipp Cimiano, Christian Chiarcos, John P. McCrae, and Jorge Gracia. 2020. *Linguistic Linked Data - Representation, Generation and Applications*. Springer.

Thierry Declerck. 2019. Ontolex as a possible bridge between wordnets and full lexical descriptions. In *Proceedings of Global WordNet Conference 2019*.

Thomas Hanke. 2004. HamNoSys – representing sign language data in language resources and language processing contexts. In *Proceedings of the LREC2004 Workshop on the Representation and Processing of Sign Languages: From SignWriting to Image Processing. Information techniques and their implications for teaching, documentation and communication*, pages 1–6, Lisbon, Portugal. European Language Resources Association (ELRA).

Maria Kopf, Marc Schulder, and Thomas Hanke. 2022. D6.1 overview of datasets for the sign languages of europe.

Jette Kristoffersen and Thomas Troelsgård. 2010. The danish sign language dictionary. In *Proceedings of the 14th EURALEX International Congress*, pages 1549–1554, Leeuwarden/Ljouwert, The Netherlands. Fryske Akademy.

Claudia Kunze and Lothar Lemnitzer. 2002. Germanet - representation, visualization, application. In *LREC*. European Language Resources Association.

Henrik Lorentzen. 2004. The danish dictionary at large: presentation, problems and perspectives. In

Proceedings of the 11th EURALEX International Congress, pages 285–294, Lorient, France. Université de Bretagne-Sud, Faculté des lettres et des sciences humaines.

Colin Lualdi, Elaine Wright, Jack Hudson, Naomi Caselli, and Christiane Fellbaum. 2021. Implementing ASLNet V1.0: Progress and Plans. In *Proceedings of the 11th Global Wordnet Conference, GWC 2021, University of South Africa (UNISA), Potchefstroom, South Africa, January 18-21, 2021*, pages 63–72. Global Wordnet Association.

Silke Matthes, Thomas Hanke, Anja Regen, Jakob Storz, Satu Worseck, Eleni Efthimiou, Athanasia-Lida Dimou, Annelies Braffort, John Glauert, and Eva Safar. 2012. Dicta-Sign -Building a Multilingual Sign Language Corpus. In *5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon. Satellite Workshop to the eighth International Conference on Language Resources and Evaluation (LREC-2012)*, Istanbul, Turkey.

John P. McCrae, Paul Buitelaar, and Philipp Cimiano. 2017. The OntoLex-Lemon Model: development and applications. In *Proc. of the 5th Biennial Conference on Electronic Lexicography (eLex)*.

Carolina Neves, Luísa Coheur, and Hugo Nicolau. 2020. HamNoSyS2SiGML: Translating HamNoSys into SiGML. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 6035–6039, Marseille, France. European Language Resources Association.

Ellen Ormel, Onno Crasborn, Els van der Kooij, Lianne van Dijken, Ellen Yassine Nauta, Jens Forster, and Daniel Stein. 2010. Glossing a multi-purpose sign language corpus. In *Proceedings of the LREC2010 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, pages 186–191, Valletta, Malta. European Language Resources Association (ELRA).

Bolette Sandford Pedersen, Manex Aguirrezabal Zabaleta, Sanni Nimb, Sussi Olsen, and Ida Rørmann Olsen. 2018. Towards a principled approach to sense clustering – a case study of wordnet and dictionary senses in danish. In *Proceedings of Global WordNet Conference 2018*. Global WordNet Association. Null ; Conference date: 08-01-2018 Through 12-01-2018.

Bolette Sandford Pedersen, Sanni Nimb, Jørg Asmussen, NicolaiHartvig Sørensen, Lars Trap-Jensen, and Henrik Lorentzen. 2009. DanNet — the challenge of compiling a wordnet for Danish by reusing a monolingual dictionary. *Language Resources and Evaluation*, 43(3):269–299.

Bolette Sandford Pedersen, Sanni Nimb, Ida Rørmann Olsen, and Sussi Olsen. 2019. Linking dannet with princeton wordnet. In *Global WordNet 2019 Proceedings, Wroclaw, Poland*, Poland. Oficyna Wydawnicza Politechniki Wroclawskiej.

Didier Schwab, Pauline Trial, Céline Vaschalde, Loïc Vial, Emmanuelle Esperanca-Rodier, and Benjamin Lecouteux. 2020. Providing semantic knowledge to a set of pictograms for people with disabilities: a set of links between WordNet and arasaac: Arasaac-WN. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 166–171, Marseille, France. European Language Resources Association.

V. Sutton. 1991. *Lessons in Sign Writing: Textbook*. Cent. for Sutton Movement Writ.

Thomas Troelsgård and Jette Kristoffersen. 2018. Improving lemmatisation consistency without a phonological description. the Danish Sign Language corpus and dictionary project. In *Proceedings of the LREC2018 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community*, pages 195–198, Miyazaki, Japan. European Language Resources Association (ELRA).