LDK 2023

A uniform RDF-based Representation of the Interlinking of Wordnets and Sign Language Data

Thierry Declerck[1], **Sam Bigeard**[2], **Dorianne Callus**[3], **Benjamin Matthews**[3], **Sussi Olsen**[4], **Loran Ripard Xuereb**[3]

[1]DFKI GmbH, Multilingual Technologies, Saarland Informatics Campus, D-66123 Saarbrücken, Germany
[2]Institute of German Sign Language and Communication of the Deaf University of Hamburg, Germany
[3]Institute of Linguistics and Language Technology, University of Malta, Malta
[4]Centre for Language Technology, NorS, University of Copenhagen, Denmark
declerck@dfki.de, sam.bigeard@uni-hamburg.de, dcall01@um.edu.mt,
saolsen@hum.ku.dk, benjamin.matthews@um.edu.mt,
contact for Loran Ripard Xuereb: benjamin.matthews@um.edu.mt

Abstract

We present ongoing and incremental work dealing with a Linked Data compliant representation of approaches using wordnets and possibly other lexical data, as representative semantic resources for the description of Spoken Language (SpL), for linking multilingual Sign Language (SL) data sets. The base for our work is given by data sets produced by the European EASIER research project, which makes use of shared IDs of the Open Multilingual Wordnet (OMW) infrastructure for linking SL glosses and basic lexical information associated with three SL data sets: British, German and Greek. We transformed the EASIER data sets onto RDF and OntoLex representations. We acted similarly with a Danish data set, which links Danish SL data and the wordnet for Danish. This transformation work was extended to other Nordic wordnets, aiming at supporting cross-lingual comparisons of Nordic SLs. We started recently work on the Maltese Sign Language Dictionary, with the challenge, that no Maltese wordnet is available for linking LSM to other SLs. The final objective of our work is to include SL data sets (and their conceptual cross-linking via wordnets, but also via other SpL lexical resources) in the Linguistic Linked Open Data cloud.

1 Introduction

Our work is pursued in the context of an initiative aiming at representing and publishing Sign Language (SL) data sets in the Linguistic Linked Data (LLOD) cloud, which is a subset of the Linked Open Data (LOD) cloud.[1] We can observe that SL data are not represented in the data sets currently included in the LLOD cloud. Also the "Overview of Datasets for the Sign Languages of Europe" published by the EASIER European project (Kopf et al., 2022)[2] does not mention any SL data set being available in a Linked Data compliant format.

We see in this a gap that needs to be bridged, as an important type of natural language is missing from the LLOD, while the motivation behind the creation of this infrastructure 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 a de facto standard for representing lexical information in RDF, the OntoLex-Lemon specifications,[5] already exist, we investigate the re-use of those specifications in order to accommodate the description and the publication of Sign Language data sets in the LLOD. Figure 1 displays the core module of OntoLex-Lemon.

A first experiment in representing SL data within RDF and OntoLex-Lemon was building on top of an approach consisting in using wordnets for interlinking British, German and Greek SL data, as originally described in Bigeard et al. (2022).[6] This approach makes use of shared IDs

[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]The data set was created in the context of the European project EASIER (https://www.project-easier.eu/). It is available at https://www.sign-lang.uni-hamburg.de/easier/sign-wordnet/index_core_synsets.html

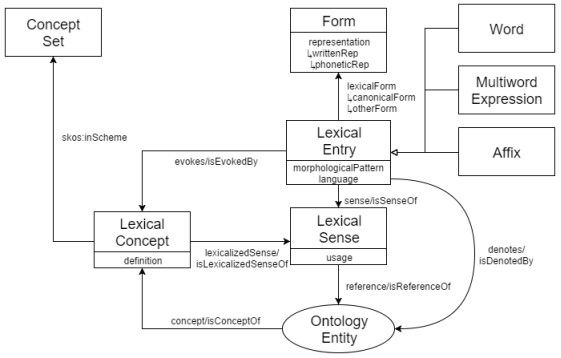


Figure 1: The core module of OntoLex-Lemon, taken from https://www.w3.org/2016/05/ontolex/

of the Open Multilingual Wordnet (OMW)[7] infrastructure as a base for interlinking SL data sets.

The OntoLex-Lemon model is also therefore a good candidate for our work, as it supports the representation of WordNet data, which are encoded with the SKOS[8] vocabulary, where the WordNet synsets are encoded as instances of the `ontolex:LexicalConcept` subclass of the `skos:Concept` class.[9] This feature offers a good starting point for transforming into RDF and OntoLex-Lemon the EASIER data sets.

Declerck et al. (2023) presents a first RDF- and OntoLex-based representation of such interlinking of OMW and SL data. Dealing with the languages covered by EASIER, adding to it French (see Section 5) and Danish (see Section 6), while starting to work also on other Nordic Languages (Declerck and Olsen, 2023).[10] We describe in this paper those stages of our incremental work, and we also introduce the most recent data set we started to work on, the Maltese Sign Language Dictionary (LSM), with a new challenge, as we cannot refer to a Maltese wordnet for cross-linking the Maltese signs to the signs of other SLs. LSM is introduced in Section 7.

[7] See (Bond and Paik, 2012) and (Bond and Foster, 2013) for more details on the Open Multilingual Wordnet and the interlinking between OMW data sets.

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

[9] See for example (Declerck, 2019).

[10] A general overview of Nordic Sign Languages is given in Bergman and Engberg-Pedersen (2010) while Aldersson and McEntee-Atalianis (2008) offer a comparison of the Icelandic and the Danish Sign Languages.

2 The Open Multilingual WordNet (OMW) Infrastructure

The motivation behind the Open Multilingual Wordnet (OMW) initiative (Bond and Paik, 2012; Bond and Foster, 2013) is to ease the use of wordnets in multiple languages. OMW proposes a shared CSV-based format for supporting the interlinking of language-specific wordnets. Version 1 of OMW[11] offers 28 wordnets,[12] all linked to the Princeton Wordnet of English (PWN),[13] which functions thus as a pivot wordnet for establishing links between all the other wordnets included in OMW (Version 1).

A very helpful feature of OMW Version 1 is given by its online search facility, where one can type a word and obtain all the related PWN synsets in user-selected languages.[14] Searching, for example, for the word "protection" we obtain 7 synsets returned. Focusing on the synset 00817680-n, with the English lemma "protection" and the Princeton WordNet gloss "the activity of protecting someone or something", we obtain the (linked) OWM lemmas for selected Nordic languages, as presented in Table 1.

Table 1: The Danish, Finnish, Norwegian (Nynorsk and Bokmål) and Swedish lemmas, linked to the shared synset ID "00817680-n", as returned by the query "protection" in the OMW search engine

Danish	forsvar, forsorg, værn, beskyttelse
Finnish	suojelu
Swedish	beskydd
Nynorsk	forsvar, beskytting, vern, omsorg
Bokmål	forsvar, beskyttelse, vern, omsorg

[11] See https://omwn.org/omw1.html

[12] While there are over 150 wordnets that have been processed by OMW, only those with a licence allowing free redistribution are listed in OMW Version 1.

[13] See (Fellbaum, 2010) for more details on WordNet. A queryable online version of PWN is available at https://wordnet.princeton.edu/

[14] https://compling.upol.cz/ntumc/cgi-bin/wn-gridx.cgi?gridmode=grid

3 Aligning several SL Resources via the Open Multilingual WordNet Infrastructure

The work reported on in this section is developed within the EASIER research project,[15] which aims to ease the communication between deaf and hearing individuals with the help of MT technologies. As such, linking different SLs through semantics is a priority. We chose to use the Open Multilingual Wordnet (OMW) infrastructure (Bond and Paik, 2012; Bond et al., 2016)[16] as a (semantic) pivot between SL data.

We are dealing with four languages (German, Greek, English and Dutch sign languages). The resources involved in our approach are the DGS corpus (Prillwitz et al., 2008), Noema+ GSL dictionary (Efthimiou et al., 2016), BSL signbank (Jordan et al., 2014), and the NGT global signbank (Crasborn et al., 2020). These resources contain various types of spoken language words associated with each sign. They may be keywords, equivalents, or SL glosses.[17] They are used as a starting point to match with the lemmas present in the corresponding aligned language versions of OMW. Then, native signers manually validate the potential matches. By using the Open Multilingual Wordnet, we aim to identify the signs with the same (or related) senses across languages.

Each resource involved has different structures, and so, the method must be flexible enough to exploit all the data available and avoid mistakes. As an example, the DGS Corpus has a multi-level structure, where each sign can be a type, a subtype, or a variant. Semantics are attached to the sub-type level. If a sense has been associated with a sub-type, it can be spread down to the variants associated with it, but not up to the type. The DGS Corpus also contains synonymy links that can be exploited to spread senses to other signs.

We describe in the following paragraphs elements of SLs that need to and could be semantically aligned across languages and language types.

Phonological transcriptions: While in an ideal world, those transcriptions from videos displaying signs could be used for establishing links between SL data for different languages, different SL data sets are transcribed with different transcription systems, e.g. HamNoSys (Hanke, 2004), SignWriting (Sutton, 2014) or others, as in the case of the Swedish SL data[18]

Besides, even if two resources use the same transcription system, the level of accuracy or precision of the transcription is not the same for all data. In some cases the transcription can be either semi-automatically generated or produced by human transcribers with different skills and views on which phonological elements of a sign should be transcribed.[19]

We are aware of efforts being made toward analysing and processing the videos directly using machine learning, rather than comparing and aligning transcriptions, but those are not in the scope of our current work.

Glosses: Many projects dealing with SL use glosses to identify signs. A gloss is, typically, a spoken language word optionally followed by a sequence of numbers or letters, to allow several signs to share the same word. The word is typically related to the meaning or iconicity of the sign, in the surrounding SpL, for easier identification. But the used word is ultimately somewhat arbitrary. Two unrelated projects working on the same sign language might have different glosses for the same sign, or the same gloss for different signs. This creates an obstacle toward linking resources together.

While many SL resources use glosses for labelling their data, the low accuracy/precision of automated tagging and the low Inter-Annotator Agreement (IAA) between human annotators for such tagging made the glosses difficult to use as a potential cross-language instrument for interlinking SL data in various languages.[20]

For linking to the IDs in OMW, we preferably use keywords and translations as a starting point to approximate the meaning of the sign, and only use glosses as a last resort. However, we use glosses as identifiers.

[15] See https://www.project-easier.eu/ for more details.

[16] See also https://omwn.org/ for more details.

[17] The term "gloss" in the SL community is carrying a different meaning as in the case of WordNet. On the specificity of glosses used for naming (or labelling) SL data in corpora, see (Ormel et al., 2010). See also further below in this section.

[18] See (Bergman and Björkstrand, 2015) for a detailed description, and also https://zrajm.github.io/teckentranskription/intro.html on recent developments on a tool to support this transcription system.

[19] Power et al. (2022), for example, report in their experiment that the similarity (but not the exact matching) of transcriptions by two undergraduate research assistants working in a related project was 0.69.

[20] Forster et al. (2010) discuss, among others, best practices for gloss annotation, in order to mitigate the issues of divergent tagging results, even in one and the same corpus.

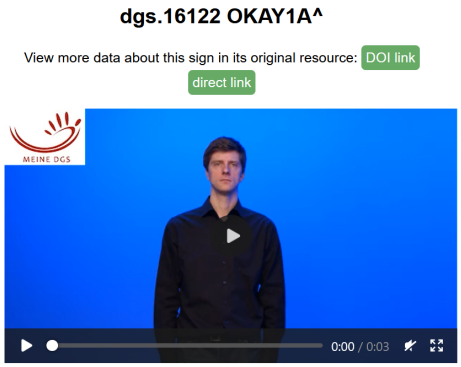
4 An Example of the Use of shared OMW IDs for interlinking SL Data

We describe in this section how the EASIER project is making use of shared OMW IDs for interlinking data in British, German and Greek Sign Languages.

omw.00806502-v approve, O.K., okay, sanction | give sanction to

- bsl.3572 goodness, virtue, good, virtuous, approve, adore, well, great, all right
- dgs.54171 $GEST-KEIN-PROBLEM1^
- dgs.13555 GUT1^
- dgs.16122 OKAY1A^
- dgs.93765 OKAY1B^
- gsl.1000 εγκρίνω

Figure 2: A screenshot showing how British, German and Greek Sign Language data are interlinked via a shared OMW index, as proposed by the EASIER project. Taken from https://www.sign-lang.uni-hamburg.de/easier/sign-wordnet/index_core_synsets.html



Synset ID and links	Synset lemmas	Synset definition	Synset examples	Type of validation	Also attested in these languages
omw.00806502-v omw link internal link	• approve • O.K. • okay • sanction	give sanction to	• I approve of his educational policies	Manual validation	BSL GSL

Figure 3: The video corresponding to the link 'dgs.16122' (see Figure 2). Taken from https://www.sign-lang.uni-hamburg.de/easier/sign-wordnet/sign/dgs.16122.html

In Figure 2, we can see that various glosses and lemmas are linked to the OMW synset omw.00806502-v. Links are directing to related videos displaying corresponding signs in three languages: BSL (British Sign Language), DGS (German Sign Language) and GSL (Greek Sign Language). Clicking on, for example, the link dgs.16122, the user is landing at the page containing the video displaying the sign, with some additional information, as shown in Figure 3.

This way, a DGS sign can be linked to both a BSL and a GSL sign, based on a shared OMW ID, which is much more accurate than going only via translation of glosses or lemmas. Those elements: videos, glosses, phonetic transcriptions (if available), links to OMW, are the elements we are encoding in a unified and harmonised Linked Data compliant format.

5 Extending the EASIER Approach with additional Signs

We searched for other SL resources in order to extend the approach described in Bigeard et al. (2022), thus linking SL data and wordnets, and then transforming those SL-wordnet combinations into RDF and OntoLex-Lemon. We found a basic lexicon of 1000 concepts associated with SL data in 4 languages, British, French, German and Greek, a result of the past Dicta-Sign project (Matthes et al., 2012), which is available at the University of Hamburg.[21] This resource is directly relevant to our purposes, as the included videos are equipped with SL glosses and HamNoSys transcriptions, as shown in Figure 4.

In Figure 4, we observe that the gloss and the HamNoSys transcription for the German video are identical with those deployed in the data used by the EASIER project for linking German SL data and wordnets, as can be seen at https://www.sign-lang.uni-hamburg.de/meinedgs/types/type13990_de.html.

This concordance of gloss and HamNoSys transcriptions[22] not only allows for the association of two videos representing this German sign to one OWM ID,[23] but it also permits the addition of signs in an additional language, French, extending

[21] https://www.sign-lang.uni-hamburg.de/dicta-sign/portal/concepts/concepts_eng.html

[22] 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.

[23] 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.

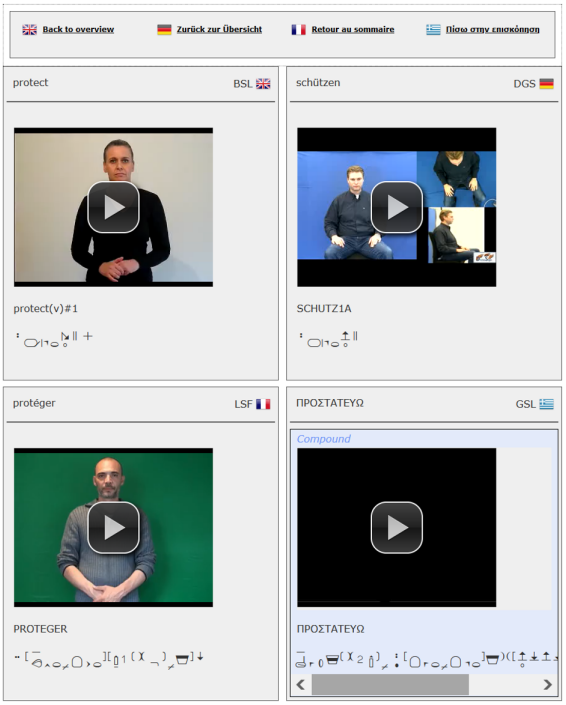


Figure 4: The concept "protect" as realised in 4 different Sign Languages. Taken from https://www.sign-lang.uni-hamburg.de/dicta-sign/portal/concepts/cs/cs_688.html

thus the multilingual coverage of the approach introduced by the EASIER project. We just need to introduce in our RDF representation new video instances (and their related glosses and transcriptions) and to link them to the same OMW ID.

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 formal representation, with the shared OMW IDs as the central component for the interlinking of the different data types and sources.

6 Extending our Work to Nordic Languages

We are extending our RDF representation to Nordic languages, while for now we have only for Danish a linking of SL data to its corresponding wordnet at our disposal.

Troelsgård and Kristoffersen (2018) discuss approaches for ensuring consistency between (Danish) Sign Language corpus data and the Dictionary of Danish signs. This approach aims at delivering a correspondence between the dictionary lemmas and the corpus lexicon, which consists of types introduced for lemmatising the tokens found in the corpus annotations (glosses added to the signs). The strategy is 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 described in Pedersen et al. (2009, 2018). This approach is thus very similar to the one described in Bigeard et al. (2022), but is monolingual. The relations between sign identifiers and lexical elements from both DanNet and other dictionary sources are encoded in a database, from which we obtained a TSV export.

In this export, we first have the signs, which correspond to entries in the Dictionary of Danish Signs (see Figure 5). A second type of data available in the export holds video links and information about the sign form (HamNoSys/SiGML).[24] A third type of information included in the export concerns the WordNet senses associated with the signs and their (form) variants.

Our work consisted thus in porting all those elements of the Danish data set to RDF and OntoLex-Lemon. 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`, thus sharing the same ID for the concept of "protection" in OMW (this holds also for French, etc.). We can therefore add the Danish sign ID (and video), which we obtained from the database, to our RDF-based infrastructure.

Figure 5: The Danish sign associated with the OMW ID "00817680-n", corresponding to the (highlighted) lemma "beskyttelse", here as one possible lexical realisation of the Danish SL gloss "FORSVARE" (*defend*)

Using the same strategy of deploying OMW as a pivot between concepts expressed in the

[24] The SiGML notation is a XML transcription of the original HamNoSys code (Neves et al., 2020)

videos, we extended our approach to Icelandic and Swedish. Through OMW we can find the lemmas for Icelandic and Swedish associated with the OMW IDs "1128193-v" and "00817680-n" (corresponding to the Danish lemmas). We use these to search in the Icelandic SignWiki,[25] and in the Swedish Sign Language Dictionary, described in Mesch et al. (2012).[26] Icelandic and Swedish glosses can be easily integrated in our RDF-based representation, as can be seen for example in Listing 1, where the gloss for the Danish sign depicted in Figure 5 is augmented with glosses or lemmas from other languages.

```
dts:GLOSS_dts-722
  rdf:type sl:GLOSS ;
  rdfs:label "\"FORSVARE\""@da ;
  rdfs:label "\"PROTEGER\""@fr ;
  rdfs:label "\"SCHUTZ1A^\""@de ;
  rdfs:label "\"protect(v)#1\""@en ;
  rdfs:label "\"beskydd\""@se ;
  rdfs:label "\"Vernda \""@is ;
.
```

Listing 1: The RDF-based representation of the gloss "FORSVARE", with the integration of multilingual labels from corresponding glosses

We further extended this approach to other Nordic languages, as described in Declerck and Olsen (2023). Data sets for 5 Nordic languages are included in OMW: Danish, Finnish, Norwegian (Nynorsk and Bokmål), and Swedish. Table 2 give some detailed information on the distribution of Nordic languages in OMW.

Table 2: Nordic wordnets included in OMW

Lang	Synsets	Words	Senses	Core
dan	4,476	4,468	5,859	81%
fin	116,763	129,839	189,227	100%
nno	3,671	3,387	4,762	66%
nob	4,455	4,186	5,586	81%
swe	6,796	5,824	6,904	99%

It is then straightforward to encode all the 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 Listing 1. The language equivalents are included, so that a Danish sign can be cross-lingually searched for, using glosses in other languages. Then, we just need to add an `ontolex:Form` instance for the Danish sign, displayed in Listing 4, and which is linked via its corresponding lexical entry to the corresponding OMW instance, shown in Figure 5.

Listing 2 shows the encoding of the Danish video already displayed in Figure 5 above, and Listing 3 shows the RDF-based representation of the corresponding gloss.

```
<http://example.org/dts#
 SignVideos_dts-722.mp4>
  rdf:type sl:SignVideos ;
  sl:hasGLOSS dts:GLOSS_dts-722 ;
  sl:hasVideoAdresss "https://www.
  tegnsprog.dk/video/t/t_2162.mp4"^^
  rdf:HTML ;
  rdfs:label "\"Video annotated with
  the gloss 'FORSVARE'\""@en ;
.
```

Listing 2: The video annotated with the gloss "FORSVARE" as an instance of the RDF class "sl:SignVideos"

```
dts:GLOSS_dts-722
  rdf:type sl:GLOSS ;
  rdfs:label "\"FORSVARE\""@da ;
.
```

Listing 3: The RDF-based representation of the gloss "FORSVARE"

Listing 4 shows a corresponding lexical form (in this case a lemma taken from OMW) and links it to the video and to the gloss it is related to, also adding the SiGML notation, which is the XML transcription of the original HamNoSys code (Neves et al., 2020).

```
dts:Form_dts-722
  rdf:type ontolex:Form ;
  sl:hasGLOSS dts:GLOSS_dts-722 ;
  sl:hasVideo <http://example.org/dts#
    SignVideos_dts-722.mp4> ;
  sl:hasVideoAdresss "https://www.
    tegnsprog.dk/video/t/t_2162.mp4"^^
    rdf:HTML ;
  rdfs:label "\"Adding transcription
    information associated with the
    video with the gloss 'FORSVARE'\""
    @en ;
  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␣;
```

[25] https://is.signwiki.org/index.php/
[26] https://teckenspraksleksikon.su.se

```
␣␣ontolex:writtenRep␣"\"beskyttelse\""
    @da␣;
.
```

Listing 4: The RDF-based representation of the lexical form related to the gloss “FORSVARE” and the corresponding video

Finally, Listing 5 displays the lexical entry for which the form is a morphological realisation. The lexical entry is pointing to the OMW ID realised as a lexical concept in OntoLex-Lemon, and which itself points to the video annotated by the one gloss.

```
dts:LexicalEntry_722
  rdf:type ontolex:LexicalEntry ;
  rdfs:label "\"forsvare, beskytte,
  beskyttelse\""@da ;
  ontolex:evokes wnid:omw-00817680-n ;
  ontolex:lexicalForm dts:Form_722 ;
.
```

Listing 5: The RDF-based representation of the lexical entry, which relates the concept and the form

```
wnid:omw-00817680-n
  rdf:type ontolex:LexicalConcept ;
  sl:hasWnLemma "\"beskydd\""@se ;
  sl:hasWnLemma "\"beskyttelse\""@da ;
  sl:hasWnLemma "\"forsorg\""@da ;
  sl:hasWnLemma "\"forsvar\""@da ;
  sl:hasWnLemma "\"protection\""@en ;
  sl:hasWnLemma "\"protection\""@fr ;
  sl:hasWnLemma "\"vernd\""@is ;
  sl:hasWnLemma "\"værn\""@da ;
  sl:hasWnLemma "\"προστασία\""@el ;

  skos:definition "\"the activity of
  protecting someone or something\""@en ;
  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 gsl:LexicalEntry_688 ;
  ontolex:isEvokedBy isl:LexicalEntry_vernda ;
  ontolex:isEvokedBy lsf:LexicalEntry_668 ;
  ontolex:isEvokedBy ssl:LexicalEntry_17861 ;
.
```

Figure 6: The encoding of the OWM ID, linking to corresponding lexical entries, which again are linked to other elements of our data set

7 The Dictionary of Maltese Sign Language (Maltese: Lingwa tas-Sinjali Maltija, LSM)

The Dizzjunarju tal-Lingwa tas-Sinjali Maltija (LSM, Maltese Sign Language) is an online dictionary comprising approximately 2,500 signs (as of 2023). Glosses for the LSM signs are in English and Maltese, so it is a trilingual dictionary. Signs are transcribed using SignWriting (Sutton, 2014), and supported by photo and video illustrations. It is not currently possible to search using the SignWriting system, but words are grouped together largely by 33 semantic categories, e.g. occupations, place names, education, travel, health, etc. This means that the dictionary may also function as a glossary for people wanting to increase vocabulary in a particular field or search for semantically related terms.

This project grew out of a linguistic corpus that was begun in 1996 at the University of Malta. It has grown well beyond this, and the original research team expanded, as well as a group of collaborators representing the wider Maltese Deaf community. The dictionary has grown through sponsorship in the form of secondments of Deaf employees working in business and government posts, as well as the hard work of Deaf and Hearing volunteers.

Maltese Sign Language is a visual-gestural language of the Maltese Deaf community. There are no official statistics available on the number of people who use LSM, though the number of people in Malta who are Deaf or Hard of Hearing is estimated to be around 1500.[27] The current form of the language is of relatively recent origin, having its sources partly in a support/play group for deaf children, which began in the mid 1970s. Malta has been an independent country since 1964, but it has maintained strong ties to the UK, and more recently to the EU. Because of the shared history, shared use of the English language, and ongoing cultural ties between the UK and Malta, there is some influence from British Sign Language (BSL) in basic signs, though the language does not appear to be part of the BSL language family. There is also influence from other signed languages. Signing systems that were used by Deaf individuals and their families before the formation of LSM in its current form are largely undocumented. Fingerspelling, a method for borrowing words from spoken languages, uses a one-handed alphabet with 29 letters of the standard Maltese alphabet. (There is a dedicated handshape for the digraph <g> but not for <ie>.)

Today, LSM classes are offered at the University of Malta, MCAST, and community settings. Significant linguistic research and documen-

[27] See http://www.deafmalta.com/ accessed: 2023-06001] for more details.

tation began in the early 2000s and has carried on (Galea, 2014; Azzopardi-Alexander, 2009, 2018; Hoffmann-Dilloway, 2021; Hoffmann-Dilloway and Xerri, 2022) The first professional interpreter began working in 2001, and Deaf interpreters have presented a daily TV news bulletin since 2012. The Maltese government passed the Maltese Sign Language Recognition Act in 2016, which provides for the promotion of the use and development of Maltese Sign Language, whilst declaring that the Maltese Sign Language is to be considered an official language of Malta. This same act also set up the Sign Language Council of Malta, which is a forum for the Deaf community to be consulted on matters relating to LSM.

There exists thus a rich dictionary for the Maltese Sign Language, but we do not have a Maltese wordnet with which we can connect the videos displaying LSM sign. We are currently working on analysing alternative semantic lexical resources, including the LSM category system, for adding a combination of Maltese SpL and SL data to our RDF-based infrastructure.

8 Conclusions and Future Work

Our RDF-based encoding results in a harmonised representation of data from both spoken and sign languages that was originally stored in different formats in different locations. Taking advantage of the work proposed by Bigeard et al. (2022) and Troelsgård and Kristoffersen (2018), we can include the links between SL data and wordnets under the umbrella of RDF and by re-using elements of OntoLex-Lemon. The Open Multilingual Wordnet infrastructure plays 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 RDF-based representation supports a dense linking of different types of information.

We are continuously extending our work to other languages. For Finnish and Norwegian we expect it to be a rather straightforward, although time consuming task, since for both these languages we have OMW entries as well as SL portals. It will be more difficult to expand to languages with fewer digital resources, as we can see while dealing with Maltese, for which we do not have a wordnet at our disposal.

The resulting data sets will be made available on Github.

Acknowledgments

The presented work is pursued in the context of the COST Action NexusLinguarum – European network for Web-centered linguistic data science (CA18209), 731015). The contributions by DFKI and University of Malta are also pursued in the context of the LT-BRIDGE project, which has received funding from the European Unions Horizon 2020 Research and Innovation Programme under Grant Agreement No 952194. Contributions by the University of Hamburg are in part supported by the EASIER (Intelligent Automatic Sign Language Translation) Project. EASIER has received funding from the European Unions Horizon 2020 research and innovation programme, grant agreement No 101016982.

References

Russell Aldersson and Lisa McEntee-Atalianis. 2008. A lexical comparison of signs from icelandic and danish sign languages. *Sign Language Studies*, 9:45–87.

Marie Azzopardi-Alexander. 2009. Iconicity and the development of maltese sign language. In Ray Fabri, editor, *Maltese Linguistics: A Snapshot in Memory of Joseph A. Cremona (1922-2003)*, pages 93–116. Brockmeyer, Bochum.

Marie Azzopardi-Alexander. 2018. Maltese Sign Language: Parallel interwoven journeys of the Deaf community and the researchers. In Patrizia Paggio and Albert Gatt, editors, *Languages of Malta*. Language Science Press.

Brita Bergman and Thomas Björkstrand. 2015. Teckentranskription. Technical Report XXV, Stockholm University, Sign Language.

Brita Bergman and Elisabeth Engberg-Pedersen. 2010. *Transmission of sign languages in the Nordic countries*, Cambridge Language Surveys, page 7494. Cambridge University Press.

Sam Bigeard, Marc Schulder, Maria Kopf, Thomas Hanke, Kyriaki Vasilaki, Anna Vacalopoulou, Theodore 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.

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 (Volume 1: Long Papers)*, pages 1352–1362, Sofia, Bulgaria. Association for Computational Linguistics.

Francis Bond and Kyonghee Paik. 2012. A survey of wordnets and their licenses. In *Proc. of the 6th Global WordNet Conference (GWC 2012)*, Matsue. 64–71.

Francis Bond, Piek Vossen, John P. McCrae, and Christiane Fellbaum. 2016. Cili: the collaborative interlingual index. In *Proceedings of the Global WordNet Conference*, volume 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.

Onno Crasborn, Richard Bank, Inge Zwitserlood, Els van der Kooij, Ellen Ormel, Johan Ros, Anique Schüller, Anne de Meijer, Merel van Zuilen, Yassine Ellen Nauta, Frouke van Winsum, and Max Vonk. 2020. Ngt dataset in global signbank.

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

Thierry Declerck and Sussi Olsen. 2023. Linked open data compliant representation of the interlinking of nordic wordnets and sign language data. In *Proceedings of the 2nd Workshop on Resources and Representations for Under-Resourced Languages and Domains*, pages 62–69.

Thierry Declerck, Thomas Troelsgård, and Sussi Olsen. 2023. Towards an rdf representation of the infrastructure consisting in using wordnets as a conceptual interlingua between multilingual sign language datasets. In *GWC 2023: 12th International Global Wordnet Conference, Proceedings*. To appear.

Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Julie A. Hochgesang, Jette Kristoffersen, and Johanna Mesch, editors. 2016. *Proceedings of the LREC2016 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining*. European Language Resources Association (ELRA), Portorož, Slovenia.

Christiane Fellbaum. 2010. Wordnet. In *Theory and applications of ontology: computer applications*, pages 231–243. Springer.

Jens Forster, Daniel Stein, Ellen Ormel, Onno Crasborn, and Hermann Ney. 2010. Best practice for sign language data collections regarding the needs of data-driven recognition and translation. In *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT 2010)*.

Maria Galea. 2014. *SignWriting (SW) of Maltese Sign Language (LSM) and its development into an orthography: Linguistic considerations*. Ph.D. thesis, University of Malta.

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).

Erika Hoffmann-Dilloway. 2021. Shadows and mirrors: Spatial and ideological perspectives on sign language competency. *Journal of Linguistic Anthropology*, 31(3):320–334.

Erika Hoffmann-Dilloway and Annabelle Xerri. 2022. # deafmum: A deaf maltese activists strategies for addressing hearing parents of deaf children. *Practicing Anthropology*, 44(4):10–14.

Fenlon Jordan, Kearsy Cormier, Ramas Rentelis, Adam Schembri, Katherine Rowley, Robert Adam, and Bencie Woll. 2014. Bsl signbank: A lexical database of british sign language (first edition).

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

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)*.

Johanna Mesch, Lars Wallin, and Thomas Björkstrand. 2012. Sign language resources in Sweden: Dictionary and corpus. In *Proceedings of the LREC2012 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon*, pages 127–130, Istanbul, Turkey. European Language Resources Association (ELRA).

Carolina Neves, Luísa Coheur, and Hugo Nicolau. 2020. HamNoSyS2SiGML: Translating HamNoSys into SiGML. In *Proceedings of the Twelfth 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, Nicolai Hartvig 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.

Justin Power, David Quinto-Pozos, and Danny Law. 2022. Signed language transcription and the creation of a cross-linguistic comparative database. In *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources*, pages 173–180, Marseille, France. European Language Resources Association.

Siegmund Prillwitz, Thomas Hanke, Susanne König, Reiner Konrad, Gabriele Langer, and Arvid Schwarz. 2008. DGS Corpus project – development of a corpus based electronic dictionary German Sign Language / German. In *Proceedings of the LREC2008 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, pages 159–164, Marrakech, Morocco. European Language Resources Association (ELRA).

V. Sutton. 2014. *Lessons in Sign Writing: Textbook*, fourth edition. Deaf Action Committee for Sign Writing.

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).