Linked Open Data compliant Representation of the Interlinking of Nordic Wordnets and Sign Language Data

Thierry Declerck
DFKI GmbH, Multilingual Technologies
Saarland Informatics Campus D3 2
D-66123 Saarbrücken, Germany
declerck@dfki.de

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

Abstract

We present ongoing work dealing with a Linked Open Data (LOD) compliant representation of Sign Language (SL) data, with the goal of supporting the cross-lingual linking of SL data, also to Spoken Language data. As the European EASIER research project has already investigated the use of Open Multilingual Wordnet (OMW) datasets for cross-linking German and Greek SL data, we propose a unified RDF-based representation of OMW and SL data. In this context, we experimented with the transformation into RDF of a rich dataset, which links Danish Sign Language data and the wordnet for Danish, DanNet. We extend this work to other Nordic languages, aiming at supporting cross-lingual comparisons of Nordic Sign Languages. This unified formal representation offers a semantic repository of information on SL data that could be accessed for supporting the creation of datasets for training or evaluating NLP applications that involve SLs.

1 Introduction

We present work that builds on top of an approach consisting in using wordnets for interlinking German and Greek Sign Language (SL) data, as described in (Bigeard et al., 2022). This approach makes use of shared IDs of the Open Multilingual Wordnet (OMW) (Bond and Paik, 2012; Bond and Foster, 2013) infrastructure as a base for interlinking the two SL datasets.

As a first step of our work dealing with the interlinking of Nordic wordnets and SL data, we investigated the work described in (Troelsgård and Kristoffersen, 2018), which is discussing the use of the Danish wordnet, DanNet (Pedersen et al., 2009, 2019), for adding relevant Danish equivalents to each sign that is already annotated with some SL glosses or words. We were happy to get access to the whole Danish dataset that lists more than 2000 sign entries, which include SL glosses, transcriptions, links to videos, associated lemmas, synonyms, English translations, DanNet IDs, etc. (see Section 4.1). We could already port a number of elements of this Danish dataset onto a Linked Data compliant representation (see Section 4.2), following our first transformation experiments, which were made with the Greek and German datasets made available by the EASIER project. [1]

We are currently extending this approach to the interlinking across Nordic SLs data, starting with the integration of the corresponding Nordic WordNet datasets, for which (Pedersen et al., 2012) gives a detailed discussion, also on the way they differ. A general overview of Nordic Sign Languages is given in (Bergman and Engberg-Pedersen, 2010), while a comparison of the Icelandic and the Danish Sign Languages is proposed in (Aldersson and McEntee-Atalianis, 2008).

Our work is anchored in the context of an initiative aiming at representing and publishing Sign Language datasets in the Linguistic Linked Data (LLOD) cloud (Cimiano et al., 2020), which is a subset of the Linked Data (LD) cloud.[2] We can observe that SL data are not represented by datasets 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)[3] we do not see any mention of a dataset being available in a Linked Data compliant format.

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

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

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

The prerequisite for publishing linguistic data in the LLOD 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 (McCrae et al., 2017),[5] already exists, we investigate as a first step the (possibly partial) re-use of those specifications in order to accommodate the description and the publication of Sign Language datasets in the LLOD. Figure 1 displays the core module of OntoLex-Lemon, which we already applied while dealing with the RDF representation of the 2 datasets made available by the EASIER project.

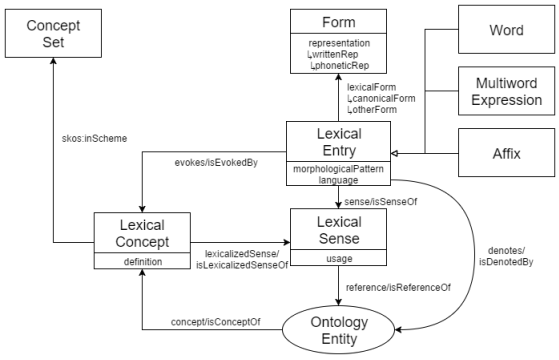

Figure 1: The core module of OntoLex-Lemon, taken from `https://www.w3.org/2016/05/ontolex/`

The OntoLex-Lemon model is also therefore a good candidate for our work, as it supports the RDF-based representation of WordNet data, which are typically encoded with the SKOS[6] vocabulary, where the WordNet synsets are encoded as instances of the `ontolex:LexicalConcept` subclass of the `skos:Concept` class.[7] This feature offers us a good starting point for transforming into RDF (and OntoLex-Lemon) the Danish dataset combining WordNet and SL data.

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 format for supporting the interlinking of language-specific wordnets. Version 1 of OMW[8] offers 28 wordnets,[9] all linked to the Princeton Wordnet of English (PWN) (Fellbaum, 2010),[10] which functions thus as a pivot wordnet for establishing links between all the other wordnets included in OMW (Version 1).

Datasets for 5 Nordic languages are included in OMW: Danish, Finnish, Norwegian (Nynorsk and Bokmål), and Swedish.[11] We also have access to the Icelandic wordnet, IceWordNet, which is stored at different sites.[12] The fact that these wordnets are (sometimes only partially) linked to the 5000 core word senses in Princeton WordNet (Boyd-Graber et al., 2006),[13] makes the interlinking of these wordnets much easier. Table 1 displays some statistics on the Nordic wordnets included in OMW, showing the proportion of their linking to the PWN core dataset.[14]

Table 1: 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%

A very helpful feature of OMW Version 1 is given by its online search facility, where one can type a word and gets all the related PWN synsets.[15] Searching, for example, for the word "protection" we get 7 synsets returned. Focusing on the synset 00817680-n, with the English lemma "protection"

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

[5] See also `https://www.w3.org/2016/05/ontolex/`

[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 `https://omwn.org/omw1.html`

[9] 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.

[10] A queryable online version of PWN is available at `https://wordnet.princeton.edu/`

[11] The Swedish wordnet included in OMW is based on SALDO (Borin et al., 2013; Borin and Forsberg, 2014).

[12] At `https://clarin.is/en/resources/iwn/` and at `https://raw.githubusercontent.com/omwn/omw-data/main/wns/isl/wn-data-isl.tab`. The data stored in GitHub is using the same OMW IDs as the other Nordic wordnets dealt with in this study.

[13] This core dataset can be accessed at `https://wordnetcode.princeton.edu/standoff-files/core-wordnet.txt`

[14] The table is derived from `https://omwn.org/omw1.html`

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

and the Princeton WordNet gloss "the activity of protecting someone or something", we get the (linked) OMW lemmas for the Nordic languages, as presented in Table 2.

Table 2: 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

3 Benefits of using OMW shared IDs for interlinking SL Data across Languages

While many SL resources are using specific SL 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 make the glosses difficult to use as a potential instrument for interlinking SL data in and across various languages. [16]

In many cases, SL data in the form of video resources are enriched with transcriptions. While in an ideal world, those transcriptions could be used for establishing links between SL data for different languages, by creating a kind of translation table between the transcription of a concept/term in one language to the transcription of the same concept/term in another language, the issue here is also that 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. (Power et al., 2022), for example, report in their experiment that the similarity (based on measuring Levenshtein distance) of transcriptions provided by two undergraduate research assistants, working in a related project, was 0.69.

Besides this point, we observe that different SL datasets are transcribed with different transcription systems, e.g. HamNoSys, (Hanke, 2004) SignWriting, (Sutton, 1991) or others, as is the case for the Swedish Sign Language data[17] or the Irish Sign Language (Murtagh et al., 2022).[18].

This issue of plurality of transcription systems makes it nearly impossible to establish a cross-linking of SL data on the basis of only those transcriptions.

4 The Danish Data and its Transformation into RDF

We describe in this section in some details first the Danish dataset we received from a designer[19] of the Dictionary of Danish Signs,[20] the building of which is described in (Kristoffersen and Troelsgård, 2010). We present then the current state of the transformation of the Danish dataset onto RDF and OntoLex-Lemon.

4.1 The Danish Dataset

Looking at the web page of the Dictionary of Danish Signs for the entry labelled with the gloss "FORSVARE", which is partly displayed in Figure 2, we observe that we find the words/lemmas "forsvar, værn, beskyttelse", among others, as potential lexical realisations of the gloss "FORSVARE" (*defend*). This leads us to hypothesise that this Danish sign corresponds to the OMW ID "00817680-n" (see Table 2). We also find verbal lexical realisations of the same sign, e.g. "forsvare, værne, beskytte", which can be attached to yet another OMW ID, namely "01128193-v, protect, 'shield from danger, injury, destruction, or damage"' (with the corresponding Danish OMW lemmas "værne, beskytte, forsvare").

The relations between sign identifiers and lexical elements from both DanNet and other dic-

[16] (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.

[17] 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.

[18] The development of this transcription system, called Sign_A is geared towards the building of fine-grained SL lexicon. More details are also given in a report of the SignON project (`https://signon-project.eu/wp-content/uploads/2022/01/SignON_D5.4_First-Sign-Language-Specific-Lexicon-and-Structure_v1.0.pdf`)

[19] We gratefully thank Thomas Troelsgård from the University College Copenhagen (KP) for giving us access to this very rich dataset.

[20] Available at `www.tegnsprog.dk`

Figure 2: 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 gloss "FORSVARE" (*defend*)

tionary sources are encoded in a database, from which we got the TSV export. In this export, looking at the entry corresponding to the page displayed in Figure 2, we first have the sign_gloss ("FORSVARE"), written in capital letters, as this is a good practice in SL datasets for indicating that we are dealing with a very generic concept or term for annotating a sign. This gloss is related to the word/lemma "beskyttelse", marked as a noun, with its English translation ("protection").

We also have the (semi-automatically) generated HamNoSys transcription, as well as its machine-readable version, using the SiGML (Neves et al., 2020) format, which is displayed in Listing 1. This type of representation can be used as the input for avatar generation (Elliott et al., 2004).

```
<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>
```

Listing 1: The SiGML code for the sign labelled with the gloss "FORSVARE".

In the TSV export we also get links to Princeton WordNet elements. Those elements can be either a so-called sense_key or a synset. For our "FORSVARE" example, we have the corresponding sense_key "protection%1:04:00::". In order to get the corresponding synset (which we need in order to establish links to other glosses in other languages), we can use the NLTK[21] method, which is displayed in Listing 2, and which for our example gives the result "`Synset('protection.n.01')`". Once we have the synset, we map it to the corresponding OMW ID. This step is currently being implemented. In case the TSV dataset includes the corresponding synset directly, we just map this one to the OMW ID, so that we can establish the comparisons with the glosses of other SL datasets, via the mediation of the OMW IDs.

```
from nltk.corpus import wordnet
    as wn
wn.synset_from_sense_key("
   protection%1:04:00::")
```

Listing 2: The NLTK code for instantiating a synset from a known sense key. Works for now only for NLTK Version 3.6.5.

Another important information included in the TSV export are the URLs pointing to the (more than 2,000) videos in which the signs are performed.

4.2 The RDF and OntoLex-Lemon Transformation of the Danish Dataset

Our work consists in porting all those (interlinked) Danish data to RDF and OntoLex-Lemon.

In the OMW version of DanNet we find, for example, the following entry "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 other wordnets in OMW. We can hereby add the Danish sign ID and video, which we got from the database, to our RDF-based infrastructure, as displayed in Listing 3 for the representation of the video and in Listing 4 for the representation of the corresponding gloss.

[21] NLTK stands for "Natural Language Toolkit", see `https://www.nltk.org/` and `https://www.nltk.org/howto/wordnet.html` for the NLTK implementation for accessing and processing WordNet data

```
<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 3: 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 4: The RDF-based representation of the gloss "FORSVARE"

It is then straightforward to establish OMW ID mediated cross-links between the signs of the various languages included in our repository, as displayed in Listing 5, showing how the Danish gloss can be enriched with labels we got from glosses extracted in our former experiment with the EASIER dataset, dealing with German and Greek. This way, we support a cross-lingual access to the Danish gloss (and the related video).

```
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 ΠΡΟΣΤΑΤΕΥΩ"\"\""@el ;
.
```

Listing 5: The RDF-based representation of the gloss "FORSVARE", with the integration of multilingual labels from corresponding glosses

Then we just have to add an `ontolex:Form` instance for the Danish sign, displayed in Listing 6, and which is linked via its corresponding lexical entry to the OMW instance (`ontolex:evokes wnid:omw-00817680-n`), as displayed in Listing 7.

```
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 ;
.
```

Listing 6: The RDF-based representation of the lexical form related to the gloss "FORSVARE" and the corresponding video

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

Listing 7: The RDF-based representation of the lexical entry, which relates the concept and the form

All the entries of the Danish dataset have been ported onto RDF and OntoLex-Lemon, already supporting a cross-lingual query of the Danish glosses and the corresponding videos.

5 A first Extension to other Nordic Sign Languages

We describe in this section first steps towards an extension of our approach to other Nordic Sign Languages. We could already start integrating a larger number of Nordic languages in our infrastructure, when considering the OMW and other sources for language specific wordnet datasets, as can be seen for the concept "potential, possible" (with OMW ID "00044353-a") in Figure 3.

Figure 3: The Nordic (and English) WordNet lemmas we could extract from the consulted sources, and encode in OntoLex-Lemon as values of an instance of the class `ontolex:LexicalConcept`

We can see in Figure 3 some differences in the way language data are associated with WordNet concepts. For the Danish case, we need to stress, that we have therefore only one lemma, as we extract the information from the Danish dataset presented in Section 4.1, and there mostly only unique senses are associated with a gloss or a OMW ID. If we would have taken the full DanNet as the source, we would have retrieved all the lemmas associated with the corresponding synset. We have now 111,166 synsets in our infrastructure, as we have been porting to RDF also the Finnish wordnet included in OMW, and which is not limited to the core WordNet.

An additional advantage of having a unified RDF and OntoLex-Lexicon of all those different wordnets, is that we can easily query the resulting triple store. We use for this SPARQL[22] queries, and Figure 4 shows a very basic query example.

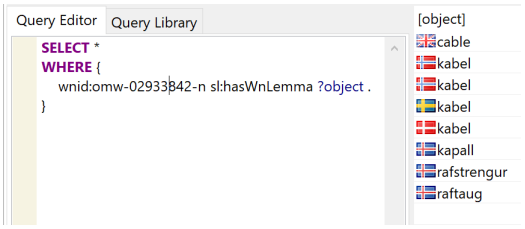

Figure 4: A very simple SPARQL query delivering the WordNet lemmas included in our RDF representation. In this example, the lemmas associated with the OMW-ID "02933842-n"

This simple SPARQL query shows how we can find the lemmas for Nordic (and other languages) associated with the OMW IDs we have already in our RDF-based infrastructure. We used those lemmas for getting a first access to glosses and videos available for the Icelandic and the Swedish SLs. For Icelandic, we consulted for now the Icelandic SignWiki, available at `https://is.signwiki.org/index.php/` and for Swedish, we consulted the Swedish Sign Language Dictionary, described in (Mesch et al., 2012), and available at `https://teckenspraklexikon.su.se`. Next step in our work will be to establish the needed cooperation with the maintainers of those sites.

6 Conclusions and Future Work

Our transformation work resulted in a harmonised representation of data from both spoken and sign languages that was originally stored in different formats in different locations and in different formats. 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 in a harmonised representation, 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 re-

[22] SPARQL is"the standard query language and protocol for Linked Open Data and RDF databases". Quoted from `https://www.ontotext.com/knowledgehub/fundamentals/what-is-sparql/`

sulting unified RDF-based representation supports a dense linking of different types of information.

We started to expand our work to other Nordic languages. We already integrated Finnish, Icelandic, Norwegian (Nynorsk and Bokmål) and Swedish in the RDF and OntoLex-Lemon representation of WordNet data. We observed that for those languages, there are SL datasets and interactive portals available. It will be more challenging to expand to Nordic languages with fewer digital resources, such as Faroese, Greenlandic and Sami, which to our knowledge have no wordnets yet, but we are aware of, for example, a portal for the Faroese SL (`https://fo.signwiki.org/index.php/B%C3%B3lkur:Or%C3%B0ab%C3%B3k_A-%C3%98`). We might be able to use the experience gained in this work as a point of departure for working on other low resourced languages, like Maltese or Slovak, as we are involved in related projects, but which are not primarily dealing with SLs.

We hope to create this way a semantically organised repository of cross-lingual (both SLs and Spoken Language data) data in the field of low-resource Sign Languages, which can be of help for supporting the creation of datasets for training or evaluating NLP applications, thinking in the first place at automated translation.

Acknowledgments

The DFKI contribution to this work is supported by the LT-BRIDGE project, which has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 952194. The work is also carried out in the context of the COST Action NexusLinguarum – European network for Web-centered linguistic data science (CA18209).

We thank Thomas Troelsgård from the University College Copenhagen (KP) for providing the Danish data, which served as the source of our transformation work into RDF and OntoLex-Lemon.

We thank also 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, access and explanations to data developed at their institute, also in the context of the EASIER project (https://www.project-easier.eu/de/).

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.

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 74–94. Cambridge University Press.

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 *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.

Lars Borin and Markus Forsberg. 2014. Swesaurus; or, the frankenstein approach to wordnet construction. In *Proceedings of the Seventh Global WordNet Conference (GWC 2014)*.

Lars Borin, Markus Forsberg, and Lennart Lönngren. 2013. Saldo: a touch of yin to wordnet's yang. *Language resources and evaluation*, 47(4):1191–1211.

Jordan Boyd-Graber, Christiane Fellbaum, Daniel Osherson, and Robert Schapire. 2006. Adding dense, weighted connections to wordnet. In *GWC 2006: 3rd International Global WordNet Conference*.

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*.

Ralph Elliott, John Glauert, Vince Jennings, and Richard Kennaway. 2004. An overview of the

SiGML notation and SiGMLSigning software system. 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 98–104, Lisbon, Portugal. European Language Resources Association (ELRA).

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)*.

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.

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).

Irene Murtagh, Víctor Ubieto Nogales, and Josep Blat. 2022. Sign language machine translation and the sign language lexicon: A linguistically informed approach. In *Proceedings of the 15th biennial conference of the Association for Machine Translation in the Americas (Volume 1: Research Track)*, pages 240–251, Orlando, USA. Association for Machine Translation in the Americas.

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.

Bolette Sandford Pedersen, Lars Borin, Markus Forsberg, Krister Lindén, Heili Orav, and Eirikur Rögnvaldsson. 2012. Linking and validating nordic and baltic wordnets- a multilingual action in meta-nord. In *Proceedings of the 6th International Global Wordnet Conference*, pages 254–260. Null ; Conference date: 09-01-2012 Through 12-01-2012.

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.

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.

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.

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).