easier
intElligent Automatic
SIgn languagE tRanslation

D6.3 INITIAL INTERLINGUAL INDEX FOR DGS AND GSL

Revision: v1.0

Work Package	WP6
Task	T6.2
Due date	30/04/2022
Submission date	30/04/2022
Deliverable lead	Universität Hamburg
Version	1.0
DOI (latest version)	https://doi.org/10.25592/uhhfdm.10170
Authors	Sam Bigeard, Marc Schulder, Maria Kopf, Thomas Hanke (Universität Hamburg – UHH) Kiriaki Vasilaki, Anna Vacalopoulou, Theodor Goulas, Athanasia–Lida Dimou, Stavroula–Evita Fotinea, Eleni Efthimiou (ILSP – Athena)
Reviewers	Fabrizio Nunnari (DFKI), Hope E. Morgan (RU)

Abstract	The purpose of the interlingual index is to link the lexical resources of all sign languages of the project. D6.3 is the first version of this index, covering DGS and GSL.
Keywords	multilingual wordnet, lexical resource, crosslingual resource, semi-automatic resource creation

Grant Agreement No.: 101016982
Call: H2020-ICT-2020-2
Topic: ICT-57-2020
Type of action: RIA

Document Revision History

Version	Date	Description of change	List of contributors
V0.1	29/04/2022	Draft for internal review	Sam Bigeard, Marc Schulder, Maria Kopf, Thomas Hanke (UHH) Kiriaki Vasilaki, Anna Vacalopoulou, Theodor Goulas, Athanasia–Lida Dimou, Stavroula–Evita Fotinea, Eleni Efthimiou (ILSP – Athena)
V0.2	29/04/2022	Internal review	Fabrizio Nunnari (DFKI), Hope E. Morgan (RU)
V1.0	30/04/2022	Camera-ready submission DOI: 10.25592/uhhfdm.10171	Sam Bigeard, Marc Schulder, Maria Kopf, Thomas Hanke (UHH) Kiriaki Vasilaki, Anna Vacalopoulou, Theodor Goulas, Athanasia–Lida Dimou, Stavroula–Evita Fotinea, Eleni Efthimiou (ILSP – Athena)

DISCLAIMER

COPYRIGHT NOTICE

Project co-funded by the European Commission in the H2020 Programme		
Nature of the deliverable	**OTHER**	
Dissemination Level		
PU	Public, fully open, e. g., web	✓
CL	Classified, information as referred to in Commission Decision 2001/844/EC	
CO	Confidential to EASIER project and Commission Services	

R: Document, report (excluding the periodic and final reports)

DEM: Demonstrator, pilot, prototype, plan designs

DEC: Websites, patents filing, press & media actions, videos, etc.

* OTHER: Software, technical diagram, etc

EXECUTIVE SUMMARY

The purpose of the inter-lingual index is to link the lexical resources from the different languages of the project and make them machine-readable. Deliverable D6.3 is the first version of this index. It includes German Sign Language (DGS) and Greek Sign Language (GSL).

The deliverable is the index itself. This report provides background on wordnet research, explains our method and choices, and presents the resulting dataset.

This version of the report refers to the state of the data on 30/04/2022, which is archived at https://doi.org/10.25592/uhhfdm.10169. For the latest data release, see https://doi.org/10.25592/uhhfdm.10168. The most up-to-date version of this report can always be found at https://doi.org/10.25592/uhhfdm.10170.

Our inter-lingual index uses the wordnet concept of synonym sets (synsets), which define concepts by gathering signs and words that can represent that meaning. By equipping a synset with signs/words from different languages, cross-lingual semantic information is established that can be used for translation and other linguistic tasks. This approach is more resistant to translation mistakes stemming from translation pairs being only valid for certain word/sign meanings. It also provides a new way to define sign types that does not rely on approximate translations to a single spoken language word, the way glosses do.

As a basis for our index, we build on the synset inventory of Open Multilingual Wordnet (OMW). Currently, 1819 GSL signs and 2230 DGS signs have been connected to one or more synsets. This includes 278 synsets connected to both a GSL and DGS signs.

In the future, we will continue expanding GSL and DGS coverage. In deliverable D6.4 we will add the other sign languages of the project: BSL, LSF, LIS, NGT and DSGS.

Funded by the Horizon 2020
Framework Programme of the European Union

CONTENTS

Executive Summary 3

List of Figures 5

List of Tables 6

Abbreviations 7

1 Introduction 8

2 Background 9
- 2.1 Wordnets for Spoken Languages . . . 9
- 2.2 Wordnets for Sign Languages . . . 9

3 Resources 11
- 3.1 GSL Lexical Resources . . . 11
- 3.2 DGS Corpus Resources . . . 11
- 3.3 Open Multilingual Wordnet . . . 12
- 3.4 Greek WordNet . . . 12
- 3.5 GermaNet . . . 13

4 Wordnet Creation 14
- 4.1 Linking the GSL Lexical Resources . . . 14
- 4.2 Linking the DGS Corpus . . . 15
- 4.3 Cross-lingual Connections . . . 16

5 Conclusion 18

References 19

Funded by the Horizon 2020 Framework Programme of the European Union

LIST OF FIGURES

4.1 Manual validation interface listing all synsets associated with a specific DGS sign. 16
4.2 Manual validation interface for comparing GSL and DGS entries associated with the same synset. 17

Funded by the Horizon 2020
Framework Programme of the European Union

LIST OF TABLES

4.1 Statistics on the state of linking GSL and DGS vocabulary to OMW. 14

Funded by the Horizon 2020
Framework Programme of the European Union

ABBREVIATIONS

OMW	Open Multilingual Wordnet
SL	sign language

Sign Languages

ASL	American Sign Language
DGS	German Sign Language / Deutsche Gebärdensprache
DSGS	Swiss-German Sign Language / Deutschschweizer Gebärdensprache
DTS	Danish Sign Language / Dansk tegnsprog
GSL	Greek Sign Language / Ελληνική νοηματική γλώσσα (Elleniké Noematiké Glossa)
LIS	Italian Sign Language / Lingua Italiana dei Segni

Funded by the Horizon 2020 Framework Programme of the European Union

1 INTRODUCTION

The purpose of the inter-lingual index is to link the lexical resources from the different languages of the project and make them machine-readable. Index uses the wordnet concept of synonym sets (synsets), which define concepts by gathering signs and words that can represent that meaning. By equipping a synset with signs/words from different languages, cross-lingual semantic information is established that can be used for translation and other linguistic tasks. This approach is more resistant to translation mistakes stemming from translation pairs being only valid for certain word/sign meanings. It also provides a new way to define sign types that does not rely on approximate translations to a single spoken language word, the way glosses do. As a basis for our index, we build on the synset inventory of Open Multilingual Wordnet (OMW).

We present our approach and results so far. We use a combination of automatic and manual methods to bootstrap the integration of sign languages into a multilingual wordnet. We start our effort on two languages in parallel: Greek Sign Language (GSL) and German Sign Language (DGS). The two languages are very different with regard to available resources, which gives us the opportunity to test different approaches and see which works best for what kind of resource.

This version of the report refers to the state of the data on 30/04/2022, which is archived at `https://doi.org/10.25592/uhhfdm.10169`. For the latest data release, see `https://doi.org/10.25592/uhhfdm.10168`. The most up-to-date version of this report can always be found at `https://doi.org/10.25592/uhhfdm.10170`.

Funded by the Horizon 2020 Framework Programme of the European Union

2 BACKGROUND

In this section we outline the relevant background on wordnets, describing the history of spoken language wordnets (Section 2.1) and existing work on sign language wordnets (Section 2.2).

2.1 WORDNETS FOR SPOKEN LANGUAGES

The concept of a wordnet was first introduced by Miller et al. (1990) as the idea of a dictionary based on psycholinguistic principles. The new approach was that words are not organised alphabetically but in so-called synonym sets (synsets), each representing an underlying concept. The synsets are interconnected via relations. While the original Princeton Wordnet (PWN) was designed for English, wordnets for many different languages have since been created. Several efforts to interconnect these into a multilingual wordnet have been undertaken. The most prominent such resource that is still actively supported is the Open Multilingual Wordnet (OMW) (Bond and Paik, 2012).

Most wordnet projects use Princeton Wordnet as a basis to expand upon, rather than developing their own Wordnet from scratch (Bond et al., 2016). This approach is known as the *expand model*. While this creates a bias toward English, it significantly reduces the amount of work needed to create a new wordnet and connect existing ones.

While the construction of a wordnet for well resourced spoken languages is relatively straightforward, the process has to be revisited for less resourced languages. Commonly used resources like dictionaries, wikis, and others may not be available. Bosch and Griesel (2017) use the *expand model* to create a wordnet of five South African languages. One of their findings is that "similarities shared on levels such as morphology or grammar and semantics allow the language teams to learn from one another, to share and thus to fast-track the development of the individual wordnets in this way" (Bosch and Griesel, 2017, p. 11). On this basis, we expect that once a wordnet for one sign language is established, subsequent sign language wordnets will be able to build on it, significantly reducing the amount of work needed.

2.2 WORDNETS FOR SIGN LANGUAGES

Work on creating wordnets for individual signed languages has been reported for Swiss-German Sign Language (DSGS) (Ebling et al., 2012), Italian Sign Language (LIS) (Shoaib et al., 2014) and American Sign Language (ASL) (Lualdi et al., 2021), although no publicly available resources have yet been released. All of these works have in common that they seek to link wordnet structures to existing lexical resources of the respective signed language. This approach allows them to leverage existing video recordings and lexicographic information for individual signs, drastically reducing the cost of creating the wordnet. In the case of ASL, several lexical resources are used to increase the available vocabulary (Lualdi et al., 2021).

Other works do not seek to publish full signed language wordnets, but rather use existing wordnets for a spoken language as an aid to internal work. Troelsgård and Kristoffersen (2018) link entries in their lexical database of Danish Sign Language (DTS) to roughly matching synsets in DanNet. These links are used as an aid to lexicographers and to automatically determine potential synonyms. The authors stress that the wordnet senses do not necessarily correspond exactly to the sign senses. Langer and Schulder (2020) match lexical entries of the DGS Corpus (see Section 3.2) with wordnet lemmas to extract supersense categories for use in coarse semantic clustering for lexicographic work. The matching is done automatically, based on existing German translational equivalents for the signs and does not take into account word sense disambiguation.

Funded by the Horizon 2020
Framework Programme of the European Union

3 RESOURCES

Following the approach of other signed language wordnet creation efforts, we build directly on existing resources for GSL, DGS, Greek, and German.

3.1 GSL LEXICAL RESOURCES

The repository of GSL lexical resources has been collected, built, and annotated for years by the Institute for Language and Speech Processing (ILSP), of Athena Research Center. It mainly consists of the Noema+ bilingual dictionary (GSL and Modern Greek) and the underlying Polytropon parallel corpus, which provides example utterances involving specific signs. These were based on utterances from expert discussions which were then re-recorded in a studio environment and annotated to serve as a "golden" corpus open for sign language (SL) technologies research (Efthimiou et al., 2016; Efthimiou et al., 2018). These two resources comprise the most extensive reference pool for GSL to date and include more than 3,600 clauses in GSL.

The lexical database currently consists of approximately 12,000 entries and it has been annotated in its entirety on the basis of the Polytropon corpus. The construction and maintenance of the database is facilitated with the use of a dedicated web-based open environment that supports the creation and interlinking of GSL resources, namely, the SiS-Builder (Goulas et al., 2010).

As the Polytropon corpus consists of isolated utterances chosen to illustrate specific signs, the contribution relating to GSL is more lexicon-based rather than corpus-based. While this has the drawback of not providing the full context and authenticity of natural discourse, the advantage of this more controlled environment is that the correspondence between GSL sign and sense-appropriate Greek translation is more explicit.

3.2 DGS CORPUS RESOURCES

The DGS Corpus is an annotated corpus of 560 hours of natural discourse in DGS (Prillwitz et al., 2008). A subset of the corpus has been released publicly as the Public DGS Corpus (Jahn et al., 2018).

The DGS Corpus implements a type hierarchy, called 'double glossing' (Konrad et al., 2012, p. 88). Each *type* represents a distinct sign realisation. It is further subdivided into *subtypes*, each of which represents a lexicalised meaning of that sign. Glosses for types and subtypes are available in English and German.

This is an example of one type and its subtypes: `https://doi.org/10.25592/dgs.corpus-3.0-type-16890`. The type is glossed *AIRPLANE1^* and has the subtypes *AIRPLANE1* and *AIRPORT2*. Either of these three glosses may be used to indicate the same sign. When the sign is used to mean "airplane" or "airport", the respective subtype is used. For other less conventionalised or common uses, the type is used directly. In the case of *AIRPLANE1^* it is used in utterances where the sign is part of multi-sign expressions for person name "Uwe

Funded by the Horizon 2020 Framework Programme of the European Union

Schönfeld" or company name "Lufthansa".

Besides the gloss name, the meaning of the subtype is accessible through one or more *concept entries* associated with it in the lexical database of the DGS Corpus. Concept entries are written with German or English mixed-case orthography (as opposed to the all-caps glosses) and specify possible meanings. A concept entry represents a meaning or sense. If both sign and word have the same sense ambiguity, i.e. multiple meanings exist, but they are the same for both, only one concept is created. This makes DGS Corpus concepts coarser than wordnet synsets but more fine grained than glosses. This definition of *concept* is specific to the DGS Corpus. The general availability of *concept entries*, however, is a feature of the lexical database system used by the corpus. Other projects that use the same database system may define its use differently. This is in fact the case for the DTS corpus, which uses the same concept entry structures to represent the DanNet synsets (Troelsgård and Kristoffersen, 2018).

On the basis of the DGS Corpus a digital dictionary for DGS is currently being created, called DW-DGS (Müller et al., 2020). The dictionary provides more nuanced information on signs and their senses. The first pre-release entries are already published[1] and can be used to further feed the sign wordnet for DGS.

3.3 OPEN MULTILINGUAL WORDNET

We use OMW's pre-existing list of synsets. A synset corresponds to a single meaning or sense and is very fine-grained. It is identified by a numerical ID independent from any particular language. It contains in several languages: a definition, words having this meaning, and example sentences. Synsets are semantically linked, thus forming the "net" part of a wordnet. As an example, the data for the synset of an apple fruit can be found at the following address: `http://compling.hss.ntu.edu.sg/omw/cgi-bin/wn-gridx.cgi?usrname=&gridmode=grid&synset=07739125-n`. The apple tree species is a different synset, which can be found here: `http://compling.hss.ntu.edu.sg/omw/cgi-bin/wn-gridx.cgi?usrname=&gridmode=grid&synset=12633994-n`

OMW by itself is a collection of individually built wordnets that share the same identifiers and overall structure. One of them is the Greek WordNet.

3.4 GREEK WORDNET

The Greek WordNet included in OMW consists of 18,049 synsets, while the English PWN comprises 117,659 synsets. The Greek synsets were originally developed in the context of BalkaNet, a multilingual wordnet of Balkan languages (Grigoriadou et al., 2004). They were based on a series of Greek lexicons and corpora. In the course of our work we found that the entries of the Greek WordNet that we inspected mainly included glossed explanations of each lexical item with minimal, if any, usage examples.

[1] `http://dw-dgs.de`

Funded by the Horizon 2020 Framework Programme of the European Union

3.5 GERMANET

The largest wordnet for German is GermaNet (Hamp and Feldweg, 1997). It contains 151,843 synsets. While it is inspired by PWN, it was built independently from German resources. Due to licence restrictions it is not directly integrated into OMW. However, for 28,564 of its synsets a mapping to PWN exists, from which OMW identifiers can be inferred. For our mutlilingual wordnet we decided to use GermaNet and expand the connections to OMW.

Funded by the Horizon 2020
Framework Programme of the European Union

4 WORDNET CREATION

To create the multilingual sign wordnet, the process is that both teams – Athena for GSL and UHH for DGS – first work independently on their respective language with frequent exchanges regarding method and implementation.

Athena follows a high precision approach of identifying strong synset matches for entries in the GSL lexical database. They prioritise working manually on many signs by attributing few synsets per sign. This approach is outlined in Section 4.1.

UHH follows a high recall approach of automatically matching signs (more precisely, subtypes) to wordnet lemmas and then verifying these matches. They prioritise validating many potential synsets for a sign. This is described in Section 4.2.

As work progresses, lists of linked synsets are exchanged between the teams to allow them to prioritise those synsets also covered by the other group. Additional cross-lingual factors such as similarities in sign form are also considered, as described in Section 4.3. This pushes forward the progress towards a large interlingual index.

Statistics on the current progress of linking both languages to OMW are outlined in Table 4.1. However, more data might have been validated since the writing of this report. The latest version of this report, which includes statistics on the latest version of the data, is always accessible at `https://doi.org/10.25592/uhhfdm.10171`.

4.1 LINKING THE GSL LEXICAL RESOURCES

To link the GSL lexical database to the Greek part of OMW, Athena used data that is collected by a semi-automatic mapping process. As the only common element of both databases are Greek lemmas, this is done by matching the "Greek equivalent" entries of GSL signs with the Greek lemmas in OMW synsets; at the end of this process, each GSL entry whose Greek equivalent also appears in OMW is mapped to the respective OMW entry.

In the next stage of the process, these automatically generated associations are checked by deaf and hearing GSL experts for validity against the respective videos offered for each sign, resulting in a new "clean" database of wordnet synsets and their GSL equivalents. Of course, this is hardly a one-to-one connection, as a lot of false equivalents are revealed in the process.

	GSL validated	DGS candidates	DGS validated	GSL/DGS overlap
distinct synsets	4214	27,020	969	278
distinct signs	1819	11,856	2230	*n/a*
distinct links	4462	105,090	2610	*n/a*

Table 4.1: *Statistics on the state of linking GSL and DGS vocabulary to OMW, at the time of writing this report. Check* `https://doi.org/10.25592/uhhfdm.10170` *for report with the latest data.*

Funded by the Horizon 2020 Framework Programme of the European Union

For instance, the GSL entry "ανέκδοτο" has been automatically linked via the Greek equivalent to synset 07220586-n, which matches the English word "anecdote". However, this sense of the Greek equivalent does not correspond to the respective entry in the GSL database, where the word is associated with its much more frequent sense of "joke". The more fitting synset 06778102-n was not found during automatic matching because it has no Greek entry. All such instances are manually corrected by GSL experts.

The accuracy of the equivalents is tested against GSL examples that are linked to each of the lexical resource's entries to make sure that each corresponds to the correct definition (referred to as a *gloss*, but not to be confused with a sign language gloss) (Fellbaum, 1998), in Greek WordNet. A secondary way of double-checking whether the correct sense of each entry is selected is reviewing the other available language versions in OMW with which annotators are familiar, namely, English and French.

In addition to that, the Greek WordNet proves to be rather limited for the purposes of this experiment, as it comprises 18,049 synsets compared to the English data of PWN, which consists of 117,659 synsets. These numbers limit the linking process even more. To compensate for this, it was decided to extend the mapping of the GSL material to the richer English part of OMW at a second level. At the time of writing, 1819 GSL signs have been linked to 4214 wordnet synsets.

4.2 LINKING THE DGS CORPUS

To link the subtypes of the DGS Corpus to synsets, UHH uses a three-step method: automatic generation of candidate matches between synsets and subtypes, automatic verification of certain simple cases, and manual verification of all remaining cases.

Automatic matching is done between the lemmas of OMW synsets and the concept entries of DGS Corpus subtypes (see Sections 2.2 and 3.2). Both German and English are used for this, although German is preferred, as concept values are more precise in that languages. If the database provides no concept entry for a subtype, its gloss name is used as a fallback.

As mentioned in Section 2.1, UHH use GermaNet as their German wordnet resource and connect its entries to OMW through its partial mapping to PWN. If a German word is not present in GermaNet or a GermaNet synset has no connection to OMW, the English concept entry or gloss is used instead. For the case that there is no English translation in the DGS Corpus or no corresponding synset in OMW, a fallback solution of automatic translation of the German gloss to English is used.

At the time of writing, automatic candidate matches between 11,856 DGS subtypes and 27,020 synsets were found. Subtypes were associated with a mean of 8.6 synsets and a median of 2. This is a "long tail" situation, where most subtypes have very few senses, while heavily polysemous terms such as "have" or "good" (and their DGS counterparts) have 20 synsets or more associated with them. In many cases, the two synsets associated with the sign represent a basic and a figurative sense.

In a second automatic processing step, candidate synsets with a high likelihood of being the correct sense for a sign are identified and marked as provisionally validated. This automatic validation step selects subtypes which were matched with only a single synset and using strong

Funded by the Horizon 2020
Framework Programme of the European Union

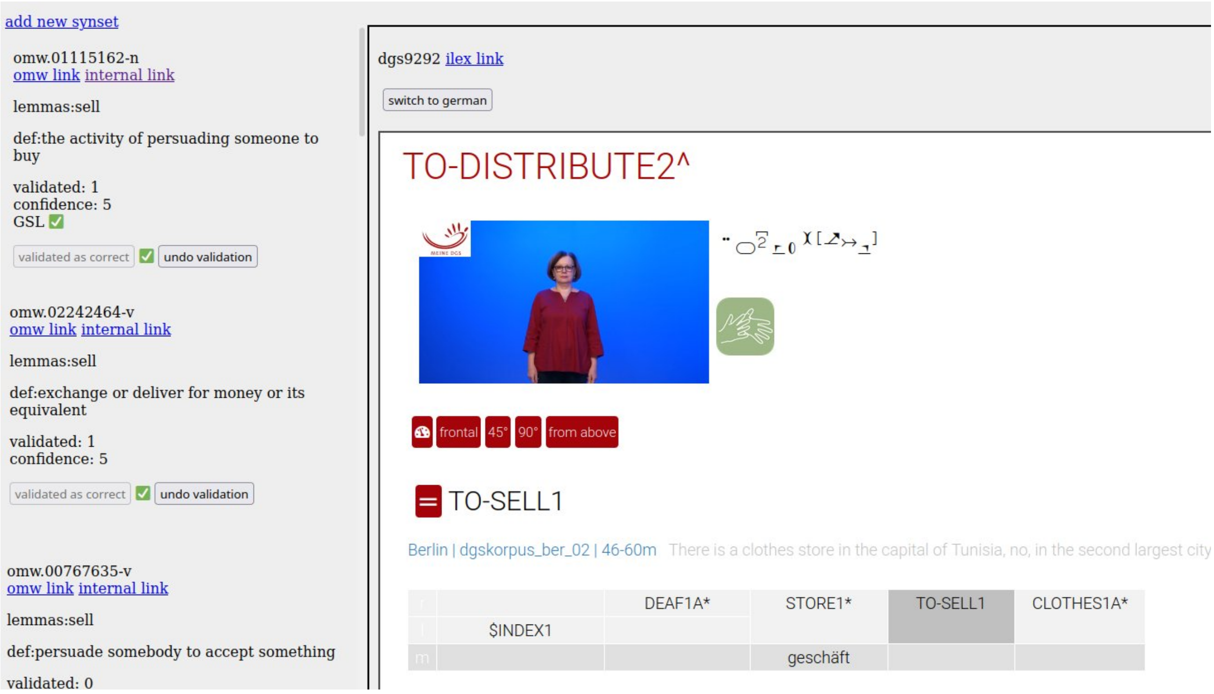

Figure 4.1: *Manual validation interface listing all synsets associated with a specific DGS sign. The left side lists the associated synsets and their validation status for DGS and GSL, while the right side shows the DGS signs' type entry page from the Public DGS Corpus website.*

match conditions, i. e. not via automatic translation. Such single match-pairs mainly occur among the long tail of homonymous expressions. As they are based on high quality human translations (concept entries or glosses), the chance of such matches introducing incorrect senses for a sign is very low.

In the final step, the remaining automatic matches are validated manually by using corpus evidence and the expert's own acceptability judgements. Ideally such verifications would only be performed by L1 language users. Due to the large number of matches (over 100,000 subtype-synset pairs) this is currently not possible for us. Instead a two-tiered approach is followed in which L2 language users validate cases for which they have high confidence and mark the remaining cases for later review by an L1 user.

This method allows us to have more annotators involved, resulting in a quicker workflow.

Figure 4.1 shows the validation interface for confirming or rejecting all synsets that were automatically matched to a specific sign. At the time of writing, 2230 DGS signs with one or more synsets have been validated.

4.3 CROSS-LINGUAL CONNECTIONS

Like other wordnet efforts for less-resourced languages, we apply the *expand model* of building on other languages already represented by a wordnet. While spoken language wordnet information is used for this out of necessity, it would be preferable to build on other sign languages

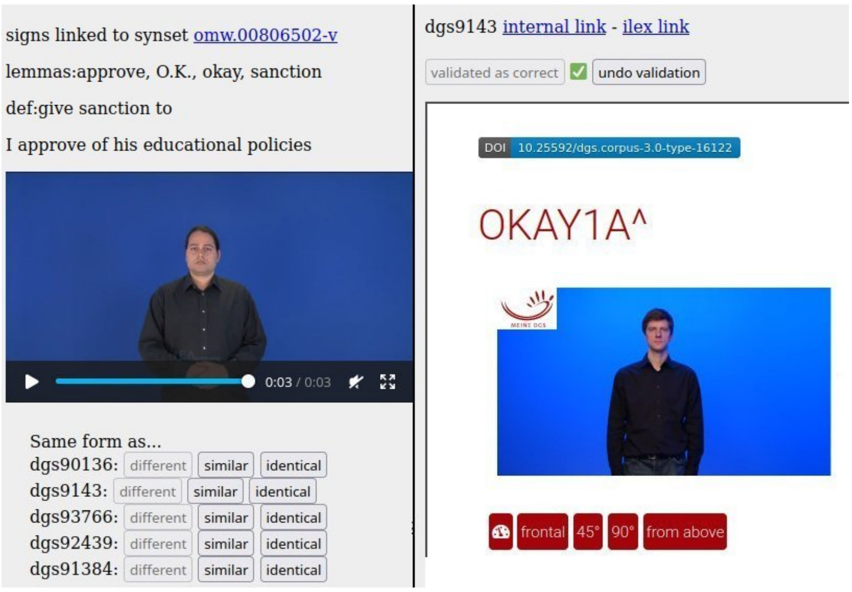

Figure 4.2: *Manual validation interface for comparing GSL and DGS entries associated with the same synset. The interface integrates GSL video from Noema+ (left) and DGS type entry pages from the Public DGS Corpus website (right). The lower left corner lists DGS types that should be compared to the given GSL video to specify whether their sign form is identical, similar or different.*

where available to be hindered less by modality-specific assumptions.

As we are working on integrating two sign languages in parallel, synergies are used where they present themselves. As Athena had already produced a number of synset-sign matches when UHH started their manual validation phase, UHH prioritised validating synsets which were covered by both automatic DGS matching and GSL.

In addition to validating synset-sign matches, the DGS also compared the form of the GSL and DGS signs (apart from mouthing) to identify identical and similar signs. The interface for this is shown in Figure 4.2. Such overlaps between languages can indicate shared iconicity (incidental or otherwise) or other kinds of linguistic relatedness. Annotating these overlaps adds a cross-lingual phonetic relation that is not usually covered by wordnets, but is of great use to research, for example for sign language technologies struggling with data sparsity.

Once signs from both languages are established for a synset, members of either team can inspect which other synsets the sign of the opposing language is connected to. They can then consider whether to expand their own sign to those synsets as well. Synsets with identical/similar forms across languages make particularly good candidates for this step.

Funded by the Horizon 2020 Framework Programme of the European Union

5 CONCLUSION

We have presented our work on connecting DGS and GSL lexical resources semantically through a multilingual wordnet. We have presented different workflows suited for lexicon-based and corpus-based data, and for cross-lingual workflows.

This work has so far resulted in a publicly available dataset of 1819 GSL signs and 2230 DGS from existing language resources being linked to 4214 and 969 OMW synsets respectively, including 278 synsets that are covered by both languages. This dataset will be regularly expanded as work progresses.

The next deliverable within this task will be D6.4, where we will add the following languages to the index: BSL, LSF, LIS, NGT, DSGS. Based on the experience of (Bosch and Griesel, 2017) with using the *expand model* for less-resourced languages, we expect the required effort for adding new languages will become progressively more manageable as other sign languages (particularly directly related ones) can be built upon.

Funded by the Horizon 2020
Framework Programme of the European Union

REFERENCES

Bond, Francis and Kyonghee Paik (2012). “A Survey of WordNets and their Licenses”. In: *Proceedings of the 6th Global WordNet Conference*. Matsue, Japan, p. 8. ISBN: 978-80-263-0244-5. URL: https://bond-lab.github.io/pdf/2012-gwc-wn-license.pdf (visited on 12/09/2021).

Bond, Francis, Piek Vossen, John Philip McCrae, and Christiane Fellbaum (2016). “CILI: the Collaborative Interlingual Index”. In: *Proceedings of the Eighth Global WordNet Conference*. Bucharest, Romania: University of Iasi, pp. 50–57. ISBN: 978-973-0-20728-6. URL: https://aclanthology.org/2016.gwc-1.9.

Bosch, Sonja E. and Marissa Griesel (2017). “Strategies for building wordnets for under-resourced languages: The case of African languages”. In: *Literator* 38.1, pp. 1–12. ISSN: 0258-2279. DOI: 10.4102/lit.v38i1.1351.

Ebling, Sarah, Katja Tissi, and Martin Volk (2012). “Semi-Automatic Annotation of Semantic Relations in a Swiss German Sign Language Lexicon”. In: *8th International Conference on Language Resources and Evaluation (LREC 2012). Proceedings of the LREC2012 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon* (Istanbul, Turkey). Ed. by Onno Crasborn, Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Jette Kristoffersen, and Johanna Mesch. Paris, France: European Language Resources Association (ELRA), pp. 31–36. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/12010.pdf.

Efthimiou, Eleni, Stavroula-Evita Fotinea, Athanasia-Lida Dimou, Theodoros Goulas, Panagiotis Karioris, Kiki Vasilaki, Anna Vacalopoulou, and Michalis Pissaris (2016). “From a Sign Lexical Database to an SL Golden Corpus – the POLYTROPON SL Resource”. In: *10th International Conference on Language Resources and Evaluation (LREC 2016). Proceedings of the LREC2016 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining* (Portorož, Slovenia). Ed. by Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Julie A. Hochgesang, Jette Kristoffersen, and Johanna Mesch. Paris, France: European Language Resources Association (ELRA), pp. 63–68. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/16003.pdf.

Efthimiou, Eleni, Kiki Vasilaki, Stavroula-Evita Fotinea, Anna Vacalopoulou, Theodoros Goulas, and Athanasia-Lida Dimou (2018). “The POLYTROPON Parallel Corpus”. In: *11th International Conference on Language Resources and Evaluation (LREC 2018). Proceedings of the LREC2018 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community* (Miyazaki, Japan). Ed. by Mayumi Bono, Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Julie A. Hochgesang, Jette Kristoffersen, Johanna Mesch, and Yutaka Osugi. Paris, France: European Language Resources Association (ELRA), pp. 39–44. ISBN: 979-10-95546-01-6. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/18043.pdf.

Fellbaum, Christiane, ed. (1998). *WordNet: an electronic lexical database*. MIT Press.

Goulas, Theodoros, Stavroula-Evita Fotinea, Eleni Efthimiou, and Michalis Pissaris (2010). “SiS-Builder: A Sign Synthesis Support Tool”. In: *7th International Conference on Language Resources and Evaluation (LREC 2010). Proceedings of the LREC2010 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies* (Valletta, Malta). Ed. by Philippe Dreuw, Eleni Efthimiou, Thomas Hanke, Trevor Johnston, Gregorio Martínez Ruiz, and Adam Schembri. Paris, France: European Language Resources Association (ELRA), pp. 102–105. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/10008.pdf.

Grigoriadou, Maria, Harry Kornilakis, Eleni Galiotou, Sofia Stamou, and Evangelos C. Papakitsos (2004). "The Software Infrastructure for the Development and Validation of the Greek Wordnet". In: *Romanian journal of information, science and technology*, pp. 89–105. URL: http://cgi.di.uoa.gr/~harryk/papers/RJIST.pdf.

Hamp, Birgit and Helmut Feldweg (1997). "GermaNet - a Lexical-Semantic Net for German". In: *Proceedings of the ACL Workshop on Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*. Madrid, Spain: Association for Computational Linguistics, pp. 9–15. URL: https://aclanthology.org/W97-0802/ (visited on 07/21/2021).

Jahn, Elena, Reiner Konrad, Gabriele Langer, Sven Wagner, and Thomas Hanke (2018). "Publishing DGS Corpus Data: Different Formats for Different Needs". In: *11th International Conference on Language Resources and Evaluation (LREC 2018). Proceedings of the LREC2018 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community* (Miyazaki, Japan). Ed. by Mayumi Bono, Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Julie A. Hochgesang, Jette Kristoffersen, Johanna Mesch, and Yutaka Osugi. Paris, France: European Language Resources Association (ELRA), pp. 83–90. ISBN: 979-10-95546-01-6. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/18018.pdf.

Konrad, Reiner, Thomas Hanke, Susanne König, Gabriele Langer, Silke Matthes, Rie Nishio, and Anja Regen (2012). "From form to function. A database approach to handle lexicon building and spotting token forms in sign languages". In: *8th International Conference on Language Resources and Evaluation (LREC 2012). Proceedings of the LREC2012 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon* (Istanbul, Turkey). Ed. by Onno Crasborn, Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Jette Kristoffersen, and Johanna Mesch. Paris, France: European Language Resources Association (ELRA), pp. 87–94. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/12023.pdf.

Langer, Gabriele and Marc Schulder (2020). "Collocations in Sign Language Lexicography: Towards Semantic Abstractions for Word Sense Discrimination". In: *12th International Conference on Language Resources and Evaluation (LREC 2020). Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives* (Marseille, France). Ed. by Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Julie A. Hochgesang, Jette Kristoffersen, and Johanna Mesch. Paris, France: European Language Resources Association (ELRA), pp. 127–134. ISBN: 979-10-95546-54-2. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/20017.pdf.

Lualdi, Colin P., Elaine Wright, Jack Hudson, Naomi K. Caselli, and Christiane Fellbaum (2021). "Implementing ASLNet V1.0: Progress and Plans". In: *Proceedings of the 11th Global Wordnet Conference*. Potchefstroom, South Africa: South African Centre for Digital Language Resources (SADiLaR), pp. 63–72. URL: https://aclanthology.org/2021.gwc-1.8 (visited on 07/21/2021).

Miller, George A., R. Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller (1990). "Introduction to WordNet: An On-line Lexical Database". In: *International Journal of Lexicography* 3. DOI: 10.1093/ijl/3.4.235.

Müller, Anke, Thomas Hanke, Reiner Konrad, Gabriele Langer, and Sabrina Wähl (2020). "From Dictionary to Corpus and Back Again – Linking Heterogeneous Language Resources for DGS". In: *12th International Conference on Language Resources and Evaluation (LREC*

Funded by the Horizon 2020 Framework Programme of the European Union

2020). Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives (Marseille, France). Ed. by Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Julie A. Hochgesang, Jette Kristoffersen, and Johanna Mesch. Paris, France: European Language Resources Association (ELRA), pp. 157–164. ISBN: 979-10-95546-54-2. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/20025.pdf.

Prillwitz, Siegmund, 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: *6th International Conference on Language Resources and Evaluation (LREC 2008). Proceedings of the LREC2008 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora* (Marrakech, Morocco). Ed. by Onno Crasborn, Eleni Efthimiou, Thomas Hanke, Ernst D. Thoutenhoofd, and Inge Zwitserlood. Paris, France: European Language Resources Association (ELRA), pp. 159–164. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/08018.pdf.

Shoaib, Umar, Nadeem Ahmad, Paolo Prinetto, and Gabriele Tiotto (2014). "Integrating MultiWordNet with Italian Sign Language lexical resources". In: *Expert Systems with Applications* 41.5, pp. 2300–2308. ISSN: 09574174. DOI: 10.1016/j.eswa.2013.09.027.

Troelsgård, Thomas and Jette Kristoffersen (2018). "Improving Lemmatisation Consistency without a Phonological Description. The Danish Sign Language Corpus and Dictionary Project." In: *11th International Conference on Language Resources and Evaluation (LREC 2018). Proceedings of the LREC2018 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community* (Miyazaki, Japan). Ed. by Mayumi Bono, Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Julie A. Hochgesang, Jette Kristoffersen, Johanna Mesch, and Yutaka Osugi. Paris, France: European Language Resources Association (ELRA), pp. 195–198. ISBN: 979-10-95546-01-6. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/18009.pdf.

Funded by the Horizon 2020 Framework Programme of the European Union