easier
intElligent Automatic
SIgn languagE tRanslation

D6.5 EXTENDED INTERLINGUAL INDEX FOR THE PROJECT'S CORE SIGN LANGUAGES AND LANGUAGES COVERED IN WORK PACKAGE 9

Revision: v1.0

Work Package	WP6
Task	T6.2
Due date	31/12/2023
Submission date	20/12/2023
Deliverable lead	Universität Hamburg (UHH)
Version	1.0
DOI (latest version)	10.25592/uhhfdm.13847
Authors	Sam Bigeard, Maria Kopf, Marc Schulder, Thomas Hanke (UHH), Kiriaki Vasilaki, Anna Vacalopoulou, Theodor Goulas, Athanasia–Lida Dimou, Stavroula–Evita Fotinea, Eleni Efthimiou (ATHENA), Neil Fox (UCL), Onno Crasborn, Lianne Westenberg (RU), Sarah Ebling, Laure Wawrinka (UZH), Johanna Mesch, Thomas Björkstrand (Stockholm University), Anna Kuder (Cologne University), Joanna Wójcicka (Warsaw University)
Reviewers	Sarah Ebling (UZH), Eleni Efthimiou (ATHENA)

Abstract	The interlingual index links sign language resources with a multilingual wordnet, connecting languages at the level of semantic concepts. This final project release of the index covers both the core sign languages of the project as well as two additional languages.

Keywords	wordnet, lexical resource, cross-lingual, 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	13/12/2023	First draft	Sam Bigeard, Maria Kopf, Marc Schulder, Thomas Hanke, Kiriaki Vasilaki, Anna Vacalopoulou, Theodor Goulas, Athanasia–Lida Dimou, Stavroula–Evita Fotinea, Eleni Efthimiou, Neil Fox, Onno Crasborn, Lianne Westenberg, Sarah Ebling, Laure Wawrinka, Johanna Mesch, Thomas Björkstrand, Anna Kuder, Joanna Wójcicka
V0.2	18/12/2023	Internal Review	Sarah Ebling (UZH), Eleni Efthimiou (ATHENA)
V1.0	20/12/2023	Camera-ready submission DOI: 10.25592/uhhfdm.13848	
V1.1	09/01/2024	Minor corrections DOI: 10.25592/uhhfdm.13910	

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

Funded by the Horizon 2020 Framework Programme of the European Union

EXECUTIVE SUMMARY

This report describes the EASIER deliverable of an interlingual index, linking signed and spoken languages at the level of individual meanings, i.e. word/sign senses. It connects different existing resources for the various languages in a machine-readable manner. This report provides background information, explains our methodology, and describes the resulting dataset.

The index has been continuously expanded throughout the project. A first version, described in EASIER deliverable D6.3, included German Sign Language (DGS) and Greek Sign Language (GSL). EASIER deliverable D6.4 introduced coverage for British Sign Language (BSL), Sign Language of the Netherlands (NGT), French Sign Language (LSF) and Swiss German Sign Language (DSGS). This deliverable introduces sign languages outside of the set of the EASIER core languages: Polish Sign Language (PJM) and Swedish Sign Language (STS). It also extends the coverage of DGS and makes data for DSGS, which so far had only been published project-internally, available to the public.

Our interlingual index uses the wordnet concept of synsets (synonym sets), which define concepts by gathering signs and words that can represent the same meaning. 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 use a three-step method: The first step is automatically matching candidate synsets to signs using the keywords and glosses associated with the sign. The second step is automatically validating links that are most likely to be correct. The final step is manual validation of the remaining links, prioritising the most useful signs.

This work has resulted in a dataset of 15,883 links, connecting 11,275 signs from eight sign languages to 11,933 synsets. The dataset as well as a web interface for browsing the data are made publicly available at https://doi.org/10.25592/dgs.wn.

Funded by the Horizon 2020
Framework Programme of the European Union

CONTENTS

Funded by the Horizon 2020
Framework Programme of the European Union

LIST OF FIGURES

Funded by the Horizon 2020
Framework Programme of the European Union

LIST OF TABLES

Funded by the Horizon 2020
Framework Programme of the European Union

ABBREVIATIONS

CILI	Collaborative InterLingual Index
OMW	Open Multilingual Wordnet
PWN	Princeton Wordnet

Sign Languages

ASL	American Sign Language
BSL	British 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
LSF	French Sign Language / Langue des Signes Française
NGT	Sign Language of the Netherlands / Nederlandse Gebarentaal
PJM	Polish Sign Language / Polski Język Migowy
STS	Swedish Sign Language / Svenskt Teckenspråk

Funded by the Horizon 2020
Framework Programme of the European Union

1 INTRODUCTION

The purpose of the interlingual index is to link lexical resources for different signed languages to each other and to spoken language information in a machine-readable manner. This deliverable follows D6.3 (Bigeard et al., 2022b) and D6.4 (Bigeard et al., 2023), which provided earlier versions of the index, covering the core project languages German Sign Language (DGS), Greek Sign Language (GSL), British Sign Language (BSL), Sign Language of the Netherlands (NGT), French Sign Language (LSF) and Swiss German Sign Language (DSGS). The version of the index presented in this deliverable extends its coverage beyond these core language to also include Swedish Sign Language (STS) and Polish Sign Language (PJM). Based on the outreach work of EASIER work package 9, inclusion of these languages was made possible through co-operations with data owners and language experts outside the project. This version also includes additional data for DGS and is the first public release of the data for DSGS, which up until now has been released only project-internally.

The 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 choosing the wrong meaning of a polysemous word/sign when deciding how to translate it. 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, but rather on (largely) language-agnostic concept representations. As a basis for our index, we build on the synset inventory of Open Multilingual Wordnet (OMW)[1] v2 (Bond et al., 2016).

This report presents our approach and the resulting dataset. We provide relevant background information in Chapter 2 and describe the lexical resources on which we build in Chapter 3. Chapter 4 describes our general methodology of combining automatic and manual annotation as well as specifics for individual languages. The resulting resource, the *Multilingual Sign Language Wordnet*, is described in Chapter 5, providing various statistics of the dataset. The report concludes with Chapter 6. The *Multilingual Sign Language Wordnet* is available as a machine-readable dataset as well as a browsable web interface. The relevant links and persistent identifiers are as follows:

- Website URL: https://sign-net.meine-dgs.de
- Website DOI: https://doi.org/10.25592/dgs.wn
- Dataset DOI: https://doi.org/10.25592/uhhfdm.13872
- Report DOI: https://doi.org/10.25592/uhhfdm.13910

[1] https://omwn.org

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. Instead of organising words by their surface form and listing them alphabetically, the researchers grouped them by so-called synonym sets, synsets for short, each representing an underlying concept. The resulting many-to-many mapping between words and synsets would describe each word as a collection of its word senses (each expressed by one synset) and each synset as a concept expressed by the set of words that could represent it. Synsets are connected to each other via various relations, most prominently among which are the hyponomy and hyperonomy relations, connecting synsets with more specific or general forms of a concept, e.g. those of wildcat, cat and feline.

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 (see Vossen, 2002). The most prominent such resource that is still actively supported is the Open Multilingual Wordnet (OMW) (Bond and Paik, 2012). It collects many openly licensed wordnets for various languages and connects them based on a shared synset inventory.

Most wordnet projects use an existing synset inventory, usually that of PWN, 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 the linguistic structures of English and anglo-centric concepts, it significantly reduces the amount of work needed to create a new wordnet and connect existing ones (Bosch and Griesel, 2017). To expand the synset inventory beyond that of PWN and introduce more language-agnostic representations, Bond et al. (2016) introduced the Collaborative InterLingual Index (CILI) and integrated it into OMW (Vossen et al., 2016). This also provides the potential to reduce the anglo-centric focus of concept definitions, a matter that will also become relevant in our efforts to introduce concepts prominent in sign languages and deaf communities.

2.2 WORDNETS FOR SIGN LANGUAGES

Work on creating wordnets for individual signed languages has previously been reported for 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 were released at the time.[2] All of these works have in common that they seek to link wordnet structures to existing lexical resources of the respective sign language. This approach allows them to

[2]The data by Ebling et al. (2012) has since been used as the basis for the DSGS component of our own multilingual sign language wordnet (cf. Bigeard et al., 2023) and will see its first (partial) public release as part of this deliverable. For more information, see Section 3.2.6.

Funded by the Horizon 2020 Framework Programme of the European Union

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 match synsets in DanNet (Pedersen et al., 2009). 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.

The DictaSign project created a cross-lingual list of 1,000 concepts for which they provided signs in four sign languages: BSL, DGS, GSL and LSF (Matthes et al., 2012). To facilitate synonym-based spoken language text search, they linked all concepts to PWN synsets (DictaSign Consortium, 2012). This was also used to provide users with text-based concept definitions (Efthimiou et al., 2012). The project's web interface, Sign-Wiki, which implemented these features, is unfortunately now defunct, but the data is available upon request.

Langer and Schulder (2020) match lexical entries of the DGS Corpus (see Section 3.2.1) 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.

Our own work on the creation of an initial multilingual sign language wordnet covering DGS and GSL has been reported in Bigeard et al. (2022a) and documented by EASIER deliverable D6.3 (Bigeard et al., 2022b). A further addition of data for BSL, NGT, LSF and DSGS is described in deliverable D6.4 (Bigeard et al., 2023). Data for each of these languages (with the exception of DSGS at the time) has been made publicly available.

Funded by the Horizon 2020 Framework Programme of the European Union

3 RESOURCES

3.1 WORDNET RESOURCES

3.1.1 Open Multilingual Wordnet

We build on the synset inventory of OMW v2 (Bond and Foster, 2013), using dataset release v1.4 (Bond and Goodman, 2021). 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, the set of words that can express this concept, and example sentences. Synsets are semantically linked, thus forming the "net" part of a wordnet. As an example, the synset `07739125-n`[3] represents an apple in the sense of the fruit. Apple in the sense of the tree species is instead represented by `12633994-n`[4], a different synset.

OMW itself is a collection of individually built wordnets from a variety of projects, connecting them through the use of shared identifiers and a unified data structure. It includes wordnets for all spoken languages of the EASIER project's set of core languages, except for German.

3.1.2 GermaNet

Since OMW does not natively support German, we need to link it to a distinct resource. The largest wordnet for German is GermaNet (Hamp and Feldweg, 1997). While it is inspired by PWN, it was built independently from German lexical resources. Due to its licence restrictions it is not directly integrated into OMW. However, for about 20% of GermaNet's synsets a mapping to PWN exists, from which OMW identifiers can be inferred. For our multilingual wordnet we decided to use GermaNet and expand the connections to OMW. We use GermaNet when working on DGS and DSGS, where our lexical resources include German words.

3.2 SIGN LANGUAGE RESOURCES

3.2.1 DGS: DGS Corpus and DW-DGS

For German Sign Language (DGS) we use the corpus and dictionary of the DGS Corpus project (Prillwitz et al., 2008). The DGS Corpus is an annotated corpus of 560 hours of natural discourse in DGS. A subset of the corpus has been released publicly as the Public DGS Corpus (Hanke et al., 2020; Konrad et al., 2020). Based on the corpus data, the project also develops DW-DGS, a digital dictionary of DGS (Müller et al., 2020).[5]

Each sign type in the DGS Corpus is associated with a German and English gloss, representing an approximate translation of the sign. The project represents signs using a hierarchical

[3] https://compling.upol.cz/ntumc/cgi-bin/wn-gridx.cgi?gridmode=grid&synset=07739125-n
[4] https://compling.upol.cz/ntumc/cgi-bin/wn-gridx.cgi?gridmode=grid&synset=12633994-n
[5] https://www.dw-dgs.de

Funded by the Horizon 2020 Framework Programme of the European Union

structure of glossed type entries, referred to as 'double glossing' (Konrad et al., 2012, p. 88). Each sign can be represented both by a supertype, which represents the distinct form of how the sign is realised, and by a subtype, which represents a specific lexicalised meaning of that sign. Each supertype can have multiple subtypes, representing its established meanings.[6]

Besides the gloss name, the meaning of a type is accessible through one or more *concept entries* associated with it in the lexical database of the DGS Corpus. Concept entries are written with regularly capitalised German or English 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 share the same set of meanings, only one concept entry is created. This makes DGS Corpus concepts coarser than wordnet synsets, as they allow some amount of polysemy, 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).

DW-DGS, while based on the same corpus data as DGS Corpus and using its annotations for lexicographic research, provides a more nuanced differentiation of sign meanings that is closer to that of synsets. Each sign receives definitions of its observed meanings. Signs that are identified as closely related variants of each other are grouped together. Further relations such as synonymy and homophony are also provided.

3.2.2 GSL: Polytropon and Noema+ dictionary

For Greek Sign Language (GSL) we rely on the repository of GSL lexical resources that 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 and lexicon (Efthimiou et al., 2016; Efthimiou et al., 2018), which provides example utterances involving specific signs. The Polytropon corpus is based on utterances from expert discussions which were then re-recorded in a studio environment and annotated to serve as a corpus open for sign language technologies research. This set of resources comprises the most extensive reference pool for GSL to date, including over 3,600 clauses and 12,000 lexical entries in GSL.

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.

[6] An example of a supertype and its subtypes: https://doi.org/10.25592/dgs.corpus-3.0-type-16890. The supertype 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" (usually accompanied by a corresponding mouthing), the respective subtype is used. For other less conventionalised or common uses, the supertype is used directly. For *AIRPLANE1^* this is the case for multi-sign expressions such as person name "Uwe Schönfeld" or company name "Lufthansa".

Funded by the Horizon 2020 Framework Programme of the European Union

3.2.3 BSL: BSL Signbank

For British Sign Language (BSL) we use the BSL Signbank (Fenlon et al., 2014)[7], which contained over 3,500 entries at time of access. The data for each sign includes video, ID gloss, and a list of English keywords. Keywords can either be synonymous or each represent a different meaning.

3.2.4 NGT: Global Signbank – NGT

For Sign Language of the Netherlands (NGT) we use the NGT dataset in Global Signbank (Crasborn et al., 2016)[8], which contained over 4,400 entries at time of access. Each sign entry provides video, ID gloss and keywords in both Dutch and English. Keywords can either be synonymous or each represent a different meaning.

3.2.5 LSF: Dicta-Sign

For French Sign Language (LSF) we use data from Dicta-Sign (Efthimiou et al., 2010)[9]. Dicta-Sign was a project that aligned 1,046 concepts to signs from BSL, DGS, LSF and GSL. The concepts were linked to PWN synsets where possible. This dataset provides links between signs and wordnet synsets, making their inclusion into the index trivial.

While the LSF part of Dicta-Sign is used for our resource, we decided against using its components for other languages, as they are all represented by other more extensive resources. It would theoretically be possible to merge the vocabularies of the different resources, but as this would require a manual comparison of signs to avoid duplicate entries, it was decided to forego this step for the time being.

3.2.6 DSGS: DSGS database

For Swiss German Sign Language (DSGS) we have access to an internal database of 3,755 signs with video, ID gloss and German keywords, kindly provided by Penny Boyes Braem. Keywords can either be synonymous or each represent a different meaning. Due to the terms of the informed consent conditions under which the videos were recorded, they cannot be made available publicly. Instead, the EASIER project partners at Zurich University are working on producing new video recordings for part of the vocabulary. As these recordings are still in production, the partners provided phonetic transcriptions in HamNoSys (Hanke, 2004) to provide a textual representation of the signs, to be replaced by videos once available.

[7] https://bslsignbank.ucl.ac.uk/
[8] https://signbank.cls.ru.nl/datasets/NGT
[9] https://www.sign-lang.uni-hamburg.de/dicta-sign/portal/

Funded by the Horizon 2020 Framework Programme of the European Union

3.2.7 STS: Swedish Sign Language Dictionary

For Swedish Sign Language (STS) we use the Swedish Sign Language Dictionary (Svenskt teckenspråkslexikon, 2023) developed at the University of Stockholm. It is a growing resource that was originally built as an online video representation of a printed dictionary (Hedberg et al., 1998) and has since been repeatedly extended, incorporating information the Swedish Sign Language Corpus (Mesch et al., 2011) and the STS language community. The dictionary covers over 20,000 sign entries at the time of writing, which contain videos, glosses, keywords in Swedish and English, meaning descriptions, and other information such as homonyms, synonyms, corpus evidence, and phonetic information.

STS was included as part of the EASIER project's outreach to include languages outside its set of core languages. As the project partners do not have language expertise in STS, all annotations were kindly provided by members of the Swedish Sign Language Dictionary (see also Section 4.1).

3.2.8 PJM: Corpus-based Dictionary of Polish Sign Language

For Polish Sign Language (PJM) we use the Corpus-based Dictionary of Polish Sign Language (CDPSL) (Łacheta et al., 2016) by the University of Warsaw. It is a dictionary building on the Corpus of Polish Sign Language (Kuder et al., 2022; Wójcicka et al., 2020), including all signs that occurred more than four times in the corpus, as well as other well-established signs. At the time of writing it contains over 6,400 sense definitions for these signs. Dictionary entries contain videos, keywords in Polish, phonetic information, meaning definitions, examples and other usage information.

PJM was included as part of the EASIER project's outreach to include languages outside its set of core languages. As the project partners do not have language expertise in PJM, all annotations were kindly provided by members of the Corpus-based Dictionary of Polish Sign Language (see also Section 4.1).

Funded by the Horizon 2020 Framework Programme of the European Union

4 WORDNET CREATION

To create wordnet content for the various sign languages, we follow the same general method for each language, combining a first phase of automatically matching sign-synset pairs and a second phase of manual validation and annotation of additional pairs. The language experts providing manual annotations for their respective language are credited in Section 4.1. The automatic matching process is described in Section 4.2, while the general workflow of manual annotation is presented in Section 4.3.

4.1 ANNOTATORS

For each language, linguistic experts with a strong grasp of the language provided annotations. Where possible, L1 signers performed this task, though due to limited resources and expert availability, fluent L2 signers were also deployed as annotators. In each case, annotators could refer to lexical information provided by the sign language resources used for our task (see Section 3.2).

The annotators are as follows:

- **DGS:** Maria Kopf
- **GSL:** Kiriaki Vasilaki, Anna Vacalopoulou, Theodor Goulas, Athanasia–Lida Dimou, Stavroula–Evita Fotinea, Eleni Efthimiou
- **BSL:** Neil Fox
- **NGT:** Onno Crasborn, Lianne Westenberg
- **DSGS:** Laure Wawrinka
- **STS:** Johanna Mesch, Thomas Björkstrand
- **PJM:** Anna Kuder, Joanna Wójcicka

In the case of LSF, all sign-synset links had already been created by the DictaSign project (Matthes et al., 2012).

4.2 AUTOMATIC MATCHING

To speed up the annotation process, we generate preliminary links between signs and synsets automatically. For this we match a sign's spoken language equivalents (glosses and keywords) to lemmas present in the pre-existing wordnet of that spoken language. Words in a contact language are preferred over English words when both are available, as the annotators who connect signs and words are usually more familiar with the contact language.

If this process results in a one-to-one match, linking a sign with only a single synset, it is marked as (tentatively) validated. As a one-to-one match indicates that the spoken language term has

Funded by the Horizon 2020 Framework Programme of the European Union

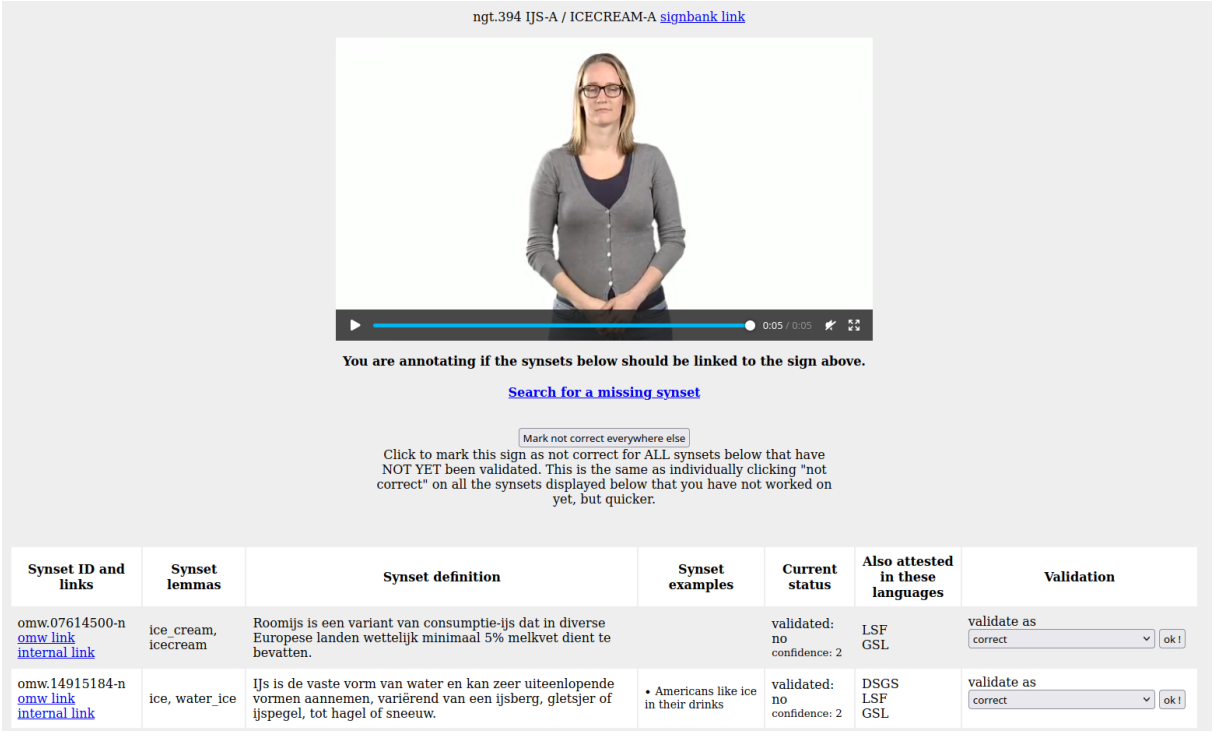


Figure 4.1: *Screenshot of the annotation interface, showing the perspective of validating all candidate synsets for a single sign. The sign's video and gloss are shown at the top, and synset candidates at the bottom. Additional synsets can be added by following the link below the video.*

only one meaning (e.g. the English word *refrigerator*) and as it has already been established as an appropriate translation of the sign, the link is assumed to very likely be correct. Automatically validated links are included in the public sign language wordnet, but marked clearly to allow users to differentiate between automatic and manual validation.

In most cases, automatic matching detects several potential sign-synset links. These are marked as possible candidates and visible to annotators, but will not be included in any dataset release until manually verified by a human annotator.

4.3 MANUAL ANNOTATION

Annotators can select perspectives showing all candidate synsets for a single sign or all signs associated with a single synset. The annotation interface is shown in Figure 4.1. Signs are represented by the video, gloss and keywords of their lexical resource. For each synset, the synset ID, contained lemmas, definition and examples are given. Where available, they are given in the annotator's preferred language, otherwise they are given in English. Links are provided to also open signs and synsets in their original resource to allow annotators to access additional information, such as corpus occurrences of a sign. For automatically matched sign-synset pairs, annotators can immediately validate them as correct or incorrect. To add additional links, annotators can open a text search interface in which they can look up signs or synsets based on their IDs, lemmas, glosses or keywords.

Funded by the Horizon 2020 Framework Programme of the European Union

We do not have the resources for an exhaustive manual annotation of each language. This leads to the question of priority. To establish connectedness not just between signed and spoken languages, but also between different sign languages, annotators were instructed to first cover synsets that were already in use and validated for other sign languages. This leads to the creation of a core set of senses with cross-lingual coverage.

Signs in the chosen language are displayed in frequency order, to prioritise the most useful signs. If the lexical database of this language is linked to a corpus, we import frequencies from it. Otherwise, we derive frequencies from the automatically matched synsets.

Annotators may be confronted with long lists of synsets per sign, some close in meaning, needing to spend extra time to understand the difference between them. This is especially the case for the most frequent signs which link to lemmas such as *to have* (20 linked synsets) or *good* (27 linked synsets). In such cases, annotators are encouraged to limit the time they spend on each sign to increase coverage, which leads to only some of the synsets of these signs being validated.

Usually wordnets only cover content word classes and do not contain function words. As one of our goals is to be able to use our resource for cross-lingual linking of languages, we forego this limitation and introduce additional 39 synsets for very common function concepts. This includes personal pronouns, interrogative pronouns, conjunctions and items specific to sign languages such as pointing and the palm-up gesture.

4.4 LANGUAGE-SPECIFIC DIFFERENCES

Depending on the resources available for a specific language, the details of the automatic and manual annotation steps must be adjusted. Among the relevant factors are the types of spoken language equivalents provided by a sign language resource (glosses, keywords, concepts), which languages these equivalents are available in, and the size, reliability and licence conditions of the relevant spoken language wordnets.

Regarding spoken language equivalents, concepts were most preferred, as they most explicitly addressed matters of synonymy. Keywords were also preferred over glosses, as they preserved the capitalisation rules of the spoken language, thus avoiding ambiguities when capitalised and lowercase forms represent distinct words or in the case of all-caps acronyms matching a normally capitalised word (e.g. the concept of *children of deaf adults* is commonly referred to as *CODA* in English, which orthographically overlaps with *coda*, the concluding passage of statement or musical composition).

Where possible, lexical information for signs was leveraged. Most commonly, polysemy was identified through the existence of multiple keywords or concepts for a sign. For DGS, the double-glossing structure of DGS Corpus was similarly used to list signs by citation form (supertype) but perform matching based on their lexicalised meanings (subtypes). With the recent introduction of DW-DGS as a resource for DGS, additional connections were made, based on synonymy links and meaning entries being shared between signs.

Our dataset follows the type structures of the individual sign language resources. Whether signs of identical form but different meaning are listed as the same or different types in our dataset depends on the structural and editorial decisions of the respective lexical resource.

Funded by the Horizon 2020
Framework Programme of the European Union

In some cases, the amount of wordnet data available for a given spoken language affected how annotation proceeded. For example, the OMW vocabulary for Greek covers “only” about 18,000 synsets, while English covers over 117,000. This caused the automatic matching step to result in more erroneous automatic validations when one-on-one word-synset matches were not due to a word having only one meaning, but rather because its set of meanings was incomplete. For example, the GSL sign for *joke* was labelled with the Greek word ανέκδοτο, which most commonly means *joke*, but in OMW is only linked to its less frequent meaning of *anecdote*. The smaller synset inventory for Greek would have also limited annotators in finding appropriate synsets, so they adjusted their workflow to also use other languages in which they were fluent for synset look-up, namely English and French. On the other hand, annotation of BSL was affected by the large synset inventory for English, which includes many nuanced differentiation. This meant that the BSL annotator tended to have to verify more synsets per sign than annotators of other languages, resulting in fewer signs being covered for a similar number of links. To focus annotation on the most relevant concepts, synsets marked by OMW as belonging to specialised domains, such as physics, were filtered out.

Funded by the Horizon 2020
Framework Programme of the European Union

5 RESULTS

This section presents the dataset created for this deliverable. Section 5.1 presents a set of examples to illustrate how the dataset was created. Section 5.2 describes the public website on which users can browse the data as well as the machine-readable data formats in which the resource is provided. Dataset statistics are provided in Section 5.3.

5.1 EXAMPLES

Figure 5.1 shows how the core sign languages of the project have been linked to the synset `omw-15236859-n`, representing the concept of the season of autumn.[10] This synset typically links to single-sense spoken language lemmas, which makes it easy to auto-validate. In our index, the synsets covered by the largest variety of languages tend to be this kind of concept.

For the given languages, the complete list of validated signs for this synset is as follows. For each sign, we show its ID, glosses, and web URL/DOI.

- `gsl.6902` φθινόπωρο video
- `lsf.56` AUTOMNE webpage
- `ngt.438` HERFST / AUTUMN webpage
- `bsl.2695` AUTUMN webpage
- `bsl.5948` AUTUMN02 webpage
- `dgs.9761` HERBST1A / AUTUMN1A doi, webpage
- `dgs.76225` HERBST1B / AUTUMN1B
- `dgs.13038` HERBST2A / AUTUMN2A
- `dgs.58031` HERBST2B / AUTUMN2B
- `dgs.74076` HERBST2C/ AUTUMN2C
- `dgs.58320` HERBST3 / AUTUMN3 doi, webpage
- `dgs.72471` HERBST4 / AUTUMN4
- `dgs.73085` HERBST5 / AUTUMN5 doi, webpage
- `dgs.74097` HERBST6A / AUTUMN6A
- `dgs.74117` HERBST6B / AUTUMN6B
- `dsgs.1354` HERBST_1B
- `dsgs.1355` HERBST_1C
- `dsgs.1356` HERBST_1E

The next example shows the potential senses of a single sign, based on the information provided by the original resources. The sign `dgs.13544` ARBEITEN1^ / TO-WORK1^[11] is linked to a number of senses. For each sense its ID, lemmas, and definition are indicated. Note that some of the correct senses are not directly linked to the lemma *work* and would not have been discovered by only comparing the sign's gloss to wordnet lemmas. This example also shows a sense that was created to fill a gap in OMW, as shown by its distinct `ea` prefix. The lemma *'what do'* is intended as a tool for the annotators to more easily find this synset.

[10] As the DSGS data we use is not public, Figure 5.1 instead shows a public stand-in provided by the Swiss Association of the Deaf, taken from https://signsuisse.sgb-fss.ch/lexikon/114810/herbst

[11] Entry in DGS Corpus: https://doi.org/10.25592/dgs.corpus-3.0-type-13544
Entry in sign wordnet: https://sign-net.meine-dgs.de/sign/dgs.13544.html

Funded by the Horizon 2020 Framework Programme of the European Union

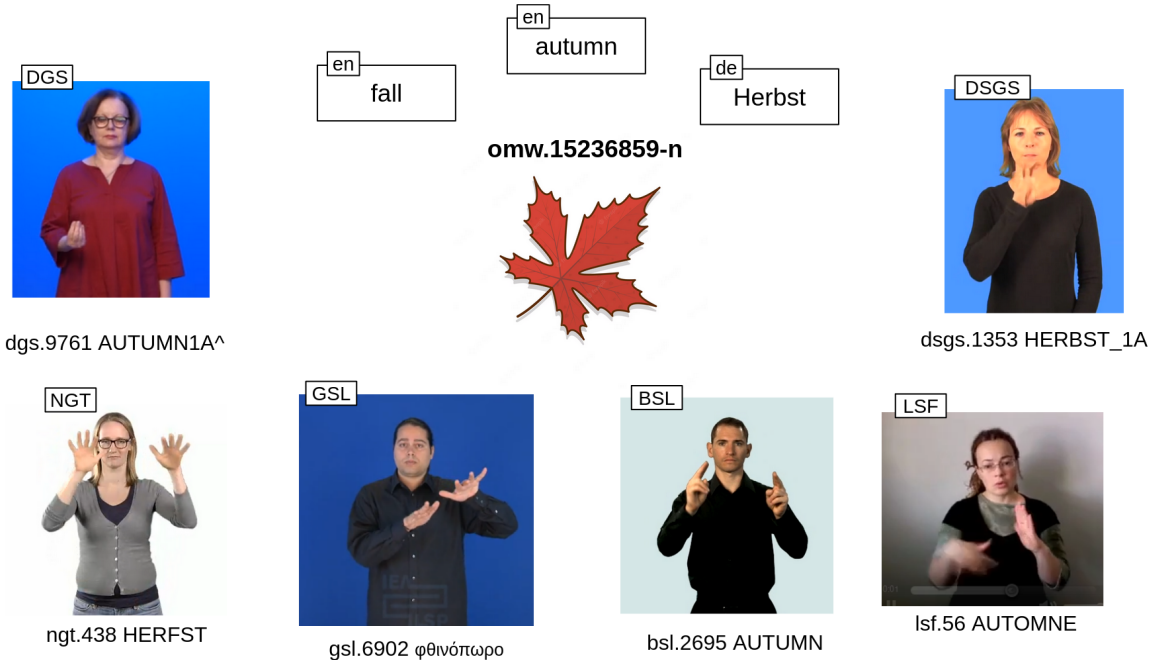


Figure 5.1: *Example of synset for autumn with some of its spoken and sign lemmas.*

The following senses are correct:

- `ea.0036` *what do*: question word, what activity a person does
- `omw.00584367-n` *employment, work*: the occupation for which you are paid
- `omw.00620752-n` *labor, labour, toil*: productive work (esp. physical work done for wages)
- `omw.02410855-v` *work, do work*: be employed
- `omw.04602044-n` *workplace, work*: a place where work is done
- `omw.13968092-n` *employment, employ*: the state of being employed or having a job
- `omw.13541167-n` *processing*: preparing or putting through a prescribed procedure

The following senses were candidates that were marked as incorrect:

- `omw.00100551-v` *exercise, work, work out*: give a workout to
- `omw.00634906-v` *solve, work out, figure out, puzzle out, lick, work*: find the solution to (a problem or question) or understand the meaning of
- `omw.01162754-v` *exploit, work*: use or manipulate to one's advantage
- `omw.01235355-v` *knead, work*: make uniform
- `omw.01659248-v` *shape, form, work, mold, mould, forge*: make something, usually for a specific function
- `omw.03841417-n` *oeuvre, work, body of work*: the total output of a writer or artist (or a substantial part of it)

Funded by the Horizon 2020 Framework Programme of the European Union

5.2 PUBLIC DATA AND WEBSITE

We release a public dataset covering the signs and synsets that have at least one verified link. Depending on the underlying lexical resource, some sign based on restricted access entries are omitted from the public release. The public dataset is licensed CC BY-NC-SA 4.0. It is available through a public web interface that allows users to browser or download the data. Figure 5.2 shows a screenshot of the page for a single sign, listing all its associated synsets, while Figure 5.3 shows the page of a synset and its signs across all covered sign languages. The website can be found at `https://sign-net.meine-dgs.de` and via the persistent identifier `https://doi.org/10.25592/dgs.wn`.

The data is also made available in machine readable formats. We provide CSV files as a general purpose file format, as well as TAB files for use with the NLTK wordnet API (Bird et al., 2009). The following files are provided:

CSV: Comma-separated tables listing dataset information. This set of files represents the most complete machine-readable form of the dataset.

- Signs: A table of all covered signs, identified by their ID in the wordnet dataset as well as their ID from the original lexical resource. For each sign its language and originating resource are given, as well as its video URL, resource URL and gloss where available.
- New synsets: The set of synsets that were newly created for this dataset. These synsets have the prefix '`ea.`'. Synsets from OMW or GermaNet are not included.
- Sign-synset links: A table of all links between signs and synsets that have been validated as correct. Each entry specifies whether validation was manual or automatic. Links that were marked as candidates but not yet validated and those that were rejected are not included.

TAB: TAB files for use with NLTK. Each file covers one language. As this formats expects lemmas to be represented by a single string, we provide alternatives for which kind of value should be represented a sign:

- Lemmas are represented by video URLs.
- Lemmas are represented by glosses.
- Lemmas are represented by the IDs used in our wordnet dataset.

Funded by the Horizon 2020 Framework Programme of the European Union

The Multilingual Sign Language Wordnet

Home | Browse BSL | Browse DGS | Browse GSL | Browse LSF | Browse NGT | Browse STS | Browse PJM | Core synsets | All synsets | Download | Credit

dgs.3603 RUND3A^

View more data about this sign in its original resource: DOI link direct link

Synset ID and links	Synset lemmas	Synset definition	Synset examples	Type of validation	Also attested in these languages
omw.00308779-n omw link internal link	• round trip	a trip to some place and back again		Automatic validation	
omw.07873807-n omw link internal link	• pizza • pizza pie	Italian open pie made of thin bread dough spread with a spiced mixture of e.g. tomato sauce and cheese		Automatic validation	DSGS
omw.06793231-n omw link internal link	• sign	a public display of a message	• he posted signs in all the shop windows	Automatic validation	GSL

The Multilingual Sign Language Wordnet v0.4
Contact | Imprint | Data Privacy

Figure 5.2: *Screenshot of the public website showing the page for one sign and its associated synsets.*

Funded by the Horizon 2020 Framework Programme of the European Union

The Multilingual Sign Language Wordnet

Home | Browse BSL | Browse DGS | Browse GSL | Browse LSF | Browse NGT | Browse STS | Browse PJM | Core synsets | All synsets | Download | Credit

Synset omw.00024356-r

View more data about this synset in its original resource: OMW link

Lemmas: no

Definition: used to express refusal or denial or disagreement etc or especially to emphasize a negative statement

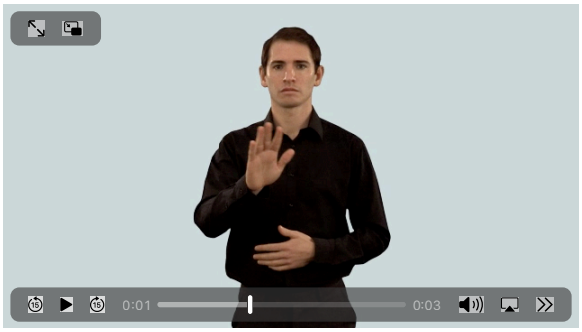
Examples:

- no, you are wrong

---

bsl.4096 don't, no, not, refuse

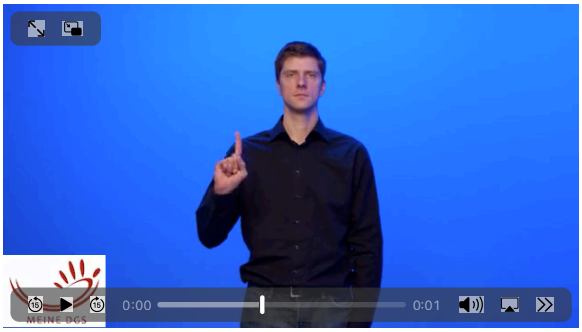
View more data about this sign in its original resource: direct link



---

dgs.15075 NEIN1A^

View more data about this sign in its original resource: DOI link direct link



---

dgs.17169 NEIN3B^

View more data about this sign in its original resource: DOI link direct link

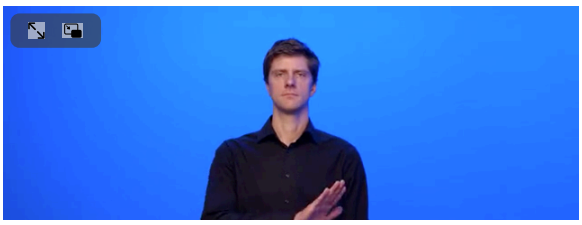


Figure 5.3: *Cropped screenshot of the public website showing the page for one synset and its associated signs from the various sign languages.*

Funded by the Horizon 2020 Framework Programme of the European Union

Published	BSL	DGS	DSGS	GSL	LSF	NGT	PJM	STS	Total
Signs	157	1604	1145	1816	1013	1500	1355	2685	11275
Synsets	449	1294	873	4167	1005	1233	965	1947	11933
Sign-Synset Links	501	2553	1184	4293	1013	1766	1802	2771	15883

Table 5.1: *The number of published signs, synsets and sign-synset links per language, including auto-validated items. For statistics only on manually validated items, see Table 5.2.*

Manually validated	BSL	DGS	DSGS	GSL	LSF	NGT	PJM	STS	Total
Signs	36	566	235	1816	1013	809	1109	87	5671
Synsets	350	579	197	4167	1005	808	728	162	7996
Sign-Synset Links	380	1043	272	4293	1013	1068	1531	173	9773

Table 5.2: *The number of published signs, synsets and sign-synset links manually validated for each language. For statistics that include auto-validated items, see Table 5.1.*

5.3 STATISTICS

The *Multilingual Sign Language Wordnet* has a total of 15,883 links, connecting 11,275 signs from eight languages to 11,933 synsets. The number of published signs, synsets and links per individual languages can be seen in Table 5.1. While automatic methods were used to assist resource creation, the involvement of language experts was essential. Annotators validated 16,663 links, marking 9,773 of them correct, thus connecting 5,671 signs and 7,996 synsets. Table 5.2 lists the number of accepted signs, synsets and links, separated by language.

As we describe in Section 4.4, differences between available sign language resources and spoken language wordnets regarding size and pre-existing information affected the validation process for individual languages. This can be seen in Table 5.3 and Figure 5.4, which show the number of sign-synset links that were automatically validated, manually accepted or manually rejected. For both LSF and GSL, only manually validated links exist, as LSF uses the pre-existing data of the Dicta-Sign project, while annotation of GSL was done internally by ATHENA team early in the EASIER project before the pipeline for the automatic candidate generation was established. Annotation for BSL focused on providing full synset coverage of highly polysemous signs, resulting in a large number of verified links for a small number of signs.

Based on feedback by RU team, the annotation interface was updated to allow users to mass reject all candidate links for a given sign or synsets, speeding up validation when systemic matching errors were encountered. RU team made good use of this feature during annotation of NGT, marking a particularly large number of incorrect links, as can be seen in Table 5.3 and Figure 5.4. We expect that similar quantities of incorrect candidates exist for the other languages and the differences in how many are manually marked are due to differing annotator workflow preferences and, in the case of earlier annotation rounds, the mass rejection feature not being introduced yet.

Our work does not only help to semantically connect signs to spoken language words. Our network of links also provides direct connections between different sign languages. 1,873 synsets are linked to two or more languages, 926 to three or more and 408 to even four or more.

Funded by the Horizon 2020 Framework Programme of the European Union

Validated links	BSL	DGS	DSGS	GSL	LSF	NGT	PJM	STS	Total
Auto: Assumed correct	223	1682	1392	0	0	698	271	2598	6864
Manual: correct	381	1048	351	4293	1013	1098	1531	173	9888
Manual: incorrect	531	898	196	0	0	3458	1126	566	6775
Total	1135	3628	1939	4293	1013	5254	2928	3337	23527

Table 5.3: *Number of validated sign-synset links by language, validation method and outcome. Automatic validation only marks candidates that are assumed to be very likely correct, while manual validation can also mark suggested candidates as incorrect.*

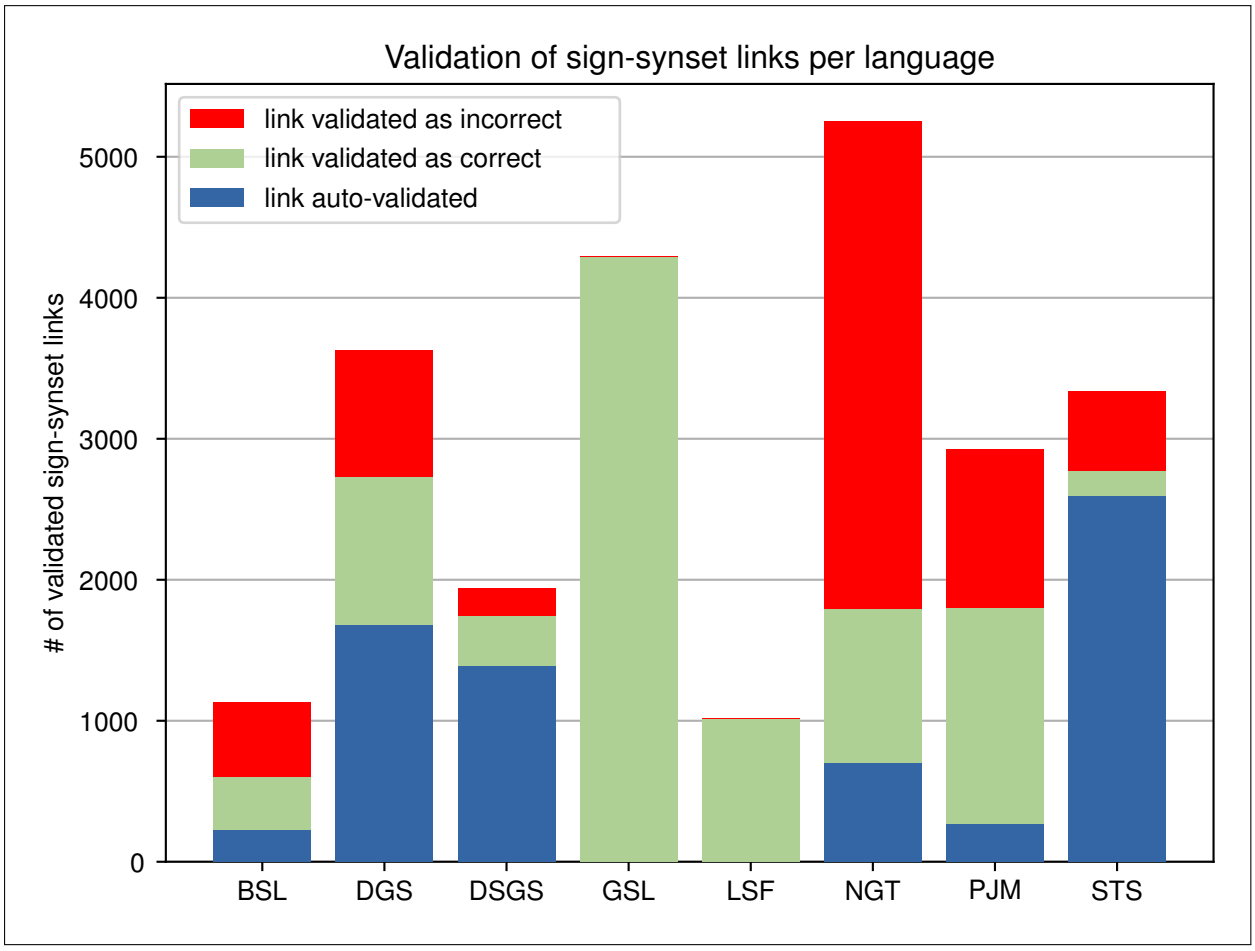


Figure 5.4: *Visual representation of Table 5.3, showing the ratio of manual to automatic validation and accepted and rejected links. Differences in validation patterns between languages are discussed in Section 5.3.*

Funded by the Horizon 2020 Framework Programme of the European Union

6 CONCLUSION

This report has presented the completed interlingual index of the EASIER project, produced in the form of a multilingual sign language wordnet. We described work from previous reports on the inclusion of DGS, GSL, BSL, NGT, LSF and DSGS as well as new work, covering the extension of data for those languages as well as the inclusion of two languages outside of the core set of project languages: STS and PJM.

The final dataset covers 15,883 links between signs and synsets, 9,773 of which were manually verified by language experts. It connects eight sign languages to the existing wordnet structures of the Open Multilingual Wordnet and GermaNet, as well as connecting them to each other at the level of semantic concepts.

The dataset has an open Creative Commons license and can be browsed and downloaded on the *Multilingual Sign Language Wordnet* website at https://sign-net.meine-dgs.de. The machine-readable data files are also archived at https://doi.org/10.25592/uhhfdm.13872.

Funded by the Horizon 2020
Framework Programme of the European Union

REFERENCES

Bigeard, Sam, Marc Schulder, Maria Kopf, Thomas Hanke, Kiki Vasilaki, Anna Vacalopoulou, Theodoros Goulas, Athanasia-Lida Dimou, Stavroula-Evita Fotinea, and Eleni Efthimiou (2022a). "Introducing Sign Languages to a Multilingual Wordnet: Bootstrapping Corpora and Lexical Resources of Greek Sign Language and German Sign Language". In: *13th International Conference on Language Resources and Evaluation (LREC 2022). Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources* (Marseille, France). Ed. by Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Julie A. Hochgesang, Jette Kristoffersen, Johanna Mesch, and Marc Schulder. Paris, France: European Language Resources Association (ELRA), pp. 9–15. ISBN: 979-10-95546-86-3. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/22036.pdf.

Bigeard, Sam, Marc Schulder, Maria Kopf, Thomas Hanke, Kiriaki Vasilaki, Anna Vacalopoulou, Theodor Goulas, Athanasia-Lida Dimou, Stavroula-Evita Fotinea, and Eleni Efthimiou (2022b). *Initial interlingual index for DGS and GSL*. Project Note D6.3. Hamburg, Germany: EASIER Consortium. DOI: 10.25592/uhhfdm.10171.

Bigeard, Sam, Marc Schulder, Maria Kopf, Thomas Hanke, Kiriaki Vasilaki, Anna Vacalopoulou, Theodor Goulas, Athanasia-Lida Dimou, Stavroula-Evita Fotinea, Eleni Efthimiou, Neil Fox, Onno Crasborn, Lianne Westenberg, Sarah Ebling, and Laure Wawrinka (2023). *Interlingual index for the project's core sign languages*. Project Note D6.4. Hamburg, Germany: EASIER Consortium. DOI: 10.25592/uhhfdm.12676.

Bird, Steven, Ewan Klein, and Edward Loper (2009). *Natural Language Processing with Python*. O'Reilly Media, Inc. 512 pp. ISBN: 978-0-596-51649-9. URL: https://www.oreilly.com/library/view/natural-language-processing/9780596803346/.

Bond, Francis and Ryan Foster (2013). "Linking and Extending an Open Multilingual Wordnet". In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*. ACL 2013. Sofia, Bulgaria: Association for Computational Linguistics, pp. 1352–1362. URL: https://aclanthology.org/P13-1133 (visited on 12/09/2021).

Bond, Francis and Michael Wayne Goodman (2021). *Open Multilingual Wordnet*. Dataset. Version 1.4. GitHub. URL: https://github.com/omwn/omw-data/releases/tag/v1.4 (visited on 12/11/2023).

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.

Crasborn, Onno, Richard Bank, Inge Zwitserlood, Els van der Kooij, Anique Schüller, Ellen Ormel, Ellen Yassine Nauta, Merel van Zuilen, Frouke van Winsum, and Johan Ros (2016). "Linking Lexical and Corpus Data for Sign Languages: NGT Signbank and the Corpus NGT". In: *10th International Conference on Language Resources and Evaluation (LREC*

Funded by the Horizon 2020 Framework Programme of the European Union

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. 41–46. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/16023.pdf.

Dicta-Sign Consortium (2012). *Sign-Wiki Prototype*. Project deliverable D6.7. Dicta-Sign Consortium. URL: https://cordis.europa.eu/docs/projects/cnect/5/231135/080/deliverables/001-D73SignWikiPrototypereportv01Ares201247395.pdf (visited on 12/12/2023).

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, Stavroula-Evita Fotinea, Thomas Hanke, John Glauert, Richard Bowden, Annelies Braffort, Christophe Collet, Petros Maragos, and François Goudenove (2010). "DICTA-SIGN: Sign Language Recognition, Generation and Modelling with application in Deaf Communication". 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. 80–83. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/10027.pdf.

Efthimiou, Eleni, Stavroula-Evita Fotinea, Thomas Hanke, John Glauert, Richard Bowden, Annelies Braffort, Christophe Collet, Petros Maragos, and François Lefebvre-Albaret (2012). "Sign Language technologies and resources of the Dicta-Sign project". 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. 37–44. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/12025.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

Funded by the Horizon 2020 Framework Programme of the European Union

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.

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)*. URL: https://bslsignbank.ucl.ac.uk/.

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).

Hanke, Thomas (2004). "HamNoSys – Representing Sign Language Data in Language Resources and Language Processing Contexts". In: *4th International Conference on Language Resources and Evaluation (LREC 2004). 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* (Lisbon, Portugal). Ed. by Oliver Streiter and Chiara Vettori. Paris, France: European Language Resources Association (ELRA), pp. 1–6. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/04001.pdf.

Hanke, Thomas, Marc Schulder, Reiner Konrad, and Elena Jahn (2020). "Extending the Public DGS Corpus in Size and Depth". 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. 75–82. ISBN: 979-10-95546-54-2. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/20016.pdf.

Hedberg, Thomas, Mikael Lindström, and Syed Ahmed Mumtaz (1998). *Svenskt teckenspråkslexikon*. Trans. by Tommy Lyxell and Elisabet Tersmeden-Lundquist. Leksand, Sweden: Sveriges dövas riksförbund (SDR). ISBN: 91-630-5542-2.

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.

Konrad, Reiner, Thomas Hanke, Gabriele Langer, Dolly Blanck, Julian Bleicken, Ilona Hofmann, Olga Jeziorski, Lutz König, Susanne König, Rie Nishio, Anja Regen, Uta Salden, Sven Wagner, Satu Worseck, and Marc Schulder (2020). *MY DGS – annotated. Public Corpus of German Sign Language, 3rd release*. Dataset. DGS-Korpus project, IDGS, Hamburg University. DOI: 10.25592/dgs.corpus-3.0.

Kuder, Anna, Joanna Wójcicka, Piotr Mostowski, and Paweł Rutkowski (2022). "Open Repository of the Polish Sign Language Corpus: Publication Project of the Polish Sign Language Corpus". In: *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources*. Marseille, France: Eu-

Funded by the Horizon 2020 Framework Programme of the European Union

ropean Language Resources Association, pp. 118–123. URL: https://aclanthology.org/2022.signlang-1.18.

Łacheta, Joanna, Małgorzata Czajkowska-Kisil, Jadwiga Linde-Usiekniewicz, and Paweł Rutkowski (2016). *Korpusowy słownik polskiego języka migowego/Corpus-based Dictionary of Polish Sign Language*. Faculty of Polish Studies, University of Warsaw. URL: https://www.slownikpjm.uw.edu.pl.

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).

Matthes, Silke, 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: *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. 117–122. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/12016.pdf.

Mesch, Johanna, Anna-Lena Nilsson, and Brita Bergman (2011). *Swedish Sign Language Corpus: Data set. The project Corpus for the Swedish Sign Language 2009-2011*. Dataset. Version 1. Department of Sign Language, Department of Linguistics, Stockholm University.

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 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.

Pedersen, Bolette S., Sanni Nimb, Jørg Asmussen, N. H. Sørensen, Lars Trap-Jensen, and Henrik Lorentzen (2009). "DanNet: the challenge of compiling a wordnet for Danish by reusing a monolingual dictionary". In: *Language Resources and Evaluation* 43, pp. 269–299. DOI: 10.1007/s10579-009-9092-1.

Funded by the Horizon 2020 Framework Programme of the European Union

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 Multi-WordNet 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.

Svenskt teckenspråkslexikon (2023). *Swedish Sign Language Dictionary online*. Lexical resource. Stockholm, Sweden: Department of Linguistics, Stockholm University. URL: https://teckensprakslexikon.su.se (visited on 06/30/2023).

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.

Vossen, Piek (2002). "WordNet, EuroWordNet and Global WordNet". In: *Revue francaise de linguistique appliquee* Vol. VII.1, pp. 27–38. ISSN: 1386-1204. DOI: 10/gntbtd.

Vossen, Piek, Francis Bond, and John McCrae (2016). "Toward a truly multilingual GlobalWordnet Grid". In: *Proceedings of the 8th Global WordNet Conference (GWC)*. GWC 2016. Ed. by Christiane Fellbaum, Piek Vossen, Verginica Barbu Mititelu, and Corina Forascu. Bucharest, Romania: Global Wordnet Association, pp. 424–431. URL: https://aclanthology.org/2016.gwc-1.59 (visited on 12/11/2023).

Wójcicka, Joanna, Anna Kuder, Piotr Mostowski, and Paweł Rutkowski, eds. (2020). *Open Repository of the Polish Sign Language Corpus*. Dataset. Warsaw, Poland: Faculty of Polish Studies, University of Warsaw. URL: https://www.korpuspjm.uw.edu.pl (visited on 06/30/2023).

Funded by the Horizon 2020 Framework Programme of the European Union