Signs and Synonymity
Continuing Development of the Multilingual Sign Language Wordnet

Marc Schulder[1], Sam Bigeard[1,2], Maria Kopf[1], Thomas Hanke[1], Anna Kuder[3], Joanna Wójcicka[4], Johanna Mesch[5], Thomas Björkstrand[5], Anna Vacalopoulou[6], Kyriaki Vasilaki[6], Theodore Goulas[6], Stavroula–Evita Fotinea[6], Eleni Efthimiou[6]

[1]Institute of German Sign Language and Communication of the Deaf, University of Hamburg, Germany
[2]Inria Centre, University of Lorraine, France
[3]Department of Linguistics, University of Cologne, Germany
[4]Department of General Linguistics, Sign Language Linguistics and Baltic Studies, University of Warsaw, Poland
[5]Department of Linguistics, Stockholm University, Sweden
[6]Institute for Language and Speech Processing, Athena Research Center, Greece
{marc.schulder, maria.kopf, thomas.hanke}@uni-hamburg.de, sam.bigeard@inria.fr, akuder@uni-koeln.de, j.filipczak@uw.edu.pl, {johanna.mesch, bjorkstrand}@ling.su.se, {avacalop, kvasilaki, tgoulas, ndimou, evita, eleni_e}@athenarc.gr

Abstract

The Multilingual Sign Language Wordnet is the first publicly available wordnet resource for sign languages. It is a growing multilingual resource providing data for eight sign languages to date. During the initial phase of its creation, the focus lay on producing the infrastructure to support various languages and to produce initial sets of content for them. This article represents the start of the second phase, in which the focus is moved to establishing overlapping coverage across the different sign languages. Building on the data produced so far, a new feature to assist annotation is introduced which leverages established partial synonymy between signs (inter- and cross-lingually) to discover likely additional synonymies. Other improvements to the annotation interface and workflow build directly on the experiences from the first phase. Working with the updated annotation interface, new data is produced for Polish Sign Language, Greek Sign Language and Swedish Sign Language.

Keywords: multilingual wordnet, sign language wordnet, resource creation, dictionary reversal

1. Introduction

The *Multilingual Sign Language Wordnet (MSL-WN)*[1] is the first publicly available wordnet resource for sign languages. It connects the inventory of several lexical sign language resources with the synset inventory of Open Multilingual Wordnet (OMW) (Bond et al., 2016). It so far provides data for eight different sign languages:

- British Sign Language (BSL)
- German Sign Language (DGS)
- Swiss German Sign Language (DSGS)
- Greek Sign Language (GSL)
- French Sign Language (LSF)
- Sign Language of the Netherlands (NGT)
- Polish Sign Language (PJM)
- Swedish Sign Language (STS)

The data of *MSL-WN* is of use both for sign language processing, e.g. to counter data sparsity issues in machine translation, and for linguistic research, especially in cross-lingual studies. For resource creators it provides an opportunity to connect different resources on a semantic level, overcoming compatibility issues caused by differences in e.g. glossing practices (Kopf et al., 2022b).

Until the end of 2023, MSL-WN was created in the context of the EU project EASIER[2]. The focus of this first phase was on establishing basic coverage for a large number of different sign languages. Lexical material for sign languages and language expertise for annotation were established through collaborations with numerous data owners. Workflows were directly informed by what data was available and what was not, leading to a number of trade-offs. Automatic assistive methods relied solely on spoken language data, requiring annotators to counter-act errors introduced by translation, differences between spoken and signed modality, and the automatic matching algorithm. Annotation for different languages also mostly worked in isola-

[1]https://doi.org/10.25592/dgs.wn

[2]https://doi.org/10.3030/101016982

Proceedings of the 11th Workshop on the Representation and Processing of Sign Languages, pages 343–353
25 May 2024.

tion from each other, as cross-lingual information was scarce, although annotators were encouraged to prioritise the verification of candidates for synsets that already had links for other languages. Nevertheless, only 16% of synsets were linked to more than one sign language.

This paper represents the start of the second phase for MSL-WN, introducing a special focus on improving cross-lingual coverage between sign languages. It is time to revisit and optimise the established workflows and to consider how the data that has been produced so far can be used to further support the annotation process. The rest of this article presents relevant related work (Section 2) and a summary of the MSL-WN creation process so far (Section 3), followed by a discussion of the lessons learned (Section 4). Section 5 introduces a new automatic suggestion feature that relies on direct (partial) synonymy between signs instead of relying on translation to a spoken language. The article is also accompanied by a new release of the MSL-WN dataset, providing new annotations for PJM, STS and GSL that were produced with the assistance of the new enhanced annotation interface (Section 6). We conclude in Section 7 with an outlook on future steps.

2. Related Work

This section discusses relevant related work regarding spoken language wordnets (Section 2.1), sign language wordnets (Section 2.2) and approaches to support the creation of lexical resources (Section 2.3).

2.1. Wordnets for Spoken Languages

The first wordnet was the Princeton Wordnet (PWN) for English (Fellbaum, 1998) and it is still the most complete and widely used wordnet in existence. It was introduced by Miller et al. (1990) as a psycholinguistically motivated alternative to the traditional approach of organising dictionaries by the alphabetical order of citation forms. Words are grouped into synsets, sets of synonyms, each of which represents a specific concept. The different senses of a polysemous word are expressed through its inclusion in several synsets. The result is a many-to-many network of forms and meanings. Furthermore, synsets are connected to each other through hyponymy and other relations, creating a taxonomic hierarchy that represents the main organisational structure of the wordnet.

Following the example of PWN, wordnets for various languages (mainly spoken languages with conventionalised written forms) have been developed since (Bond and Paik, 2012). Several projects, such as EuroWordNet (Vossen, 1998), BalkaNet (Tufiş et al., 2004) and African WordNet (Le Roux et al., 2008), have also worked on creating aligned wordnets for several languages. Many wordnets with open access licences have since been connected into an interconnected network of wordnets by the Open Multilingual Wordnet (OMW) (Bond and Paik, 2012).

2.2. Wordnets for Sign Languages

A number of reports on creating sign language wordnets exist. Work on individual sign languages was 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 the data was not made publicly available at the time. In all cases, the authors made use of existing lexical resources, allowing them to leverage available lexical information and video material to drastically reduce production cost and in turn provide added value to those lexical resources. Lualdi et al. (2021) also reported on combining several resources to increase the available vocabulary.

Other works use wordnets to support software functions or internal work processes. The DictaSign project (Matthes et al., 2012) defined a list of 1,000 concepts, each represented by a PWN synset, for which they provided signs in four languages: BSL, DGS, GSL and LSF (Dicta-Sign Consortium, 2012). The synsets were used to connect signs cross-lingually, provide concept definitions and allow synonym-based spoken language text search through the project's web interface (Efthimiou et al., 2012).

The *Danish Sign Language Corpus and Dictionary* project (Troelsgård and Kristoffersen, 2018b) link their sign type inventory to synsets from DanNet (Pedersen et al., 2009) to enhance the annotator's type search by also matching to Danish synonyms (Troelsgård and Kristoffersen, 2018a). Declerck and Olsen (2023) reported on-going work on making this information publicly available as linked open data.

Langer and Schulder (2020) automatically match lexical entries of the DGS Corpus (Prillwitz et al., 2008) with lemmas from GermaNet (Hamp and Feldweg, 1997) to extract supersense categories for use in coarse semantic clustering for lexicographic work, although no sense disambiguation is performed.

The DSGS data by Ebling et al. (2012) and the LSF data of the DictaSign project have been integrated into the MSL-WN (Bigeard et al., 2024).

2.3. Bootstrapping lexical content

Creating lexical resources is labour intensive and many methods to support or automate this work have been considered. For bilingual dictionaries, a common technique (these days aided by the use of lexicographic editing software) is that of *dictionary reversal*, in which one translation direction is produced first and then a first draft of the inverse direction is created by reversing the entries (cf. Martin, 2013). As Martin (2013) points out, relying purely on the surface forms of words for reversal can produce many mistranslations caused by incorrectly mapping the different meanings of polysemous words across languages. More robust reversals can be achieved when building on lexical units that represent individual meanings (naturally, additional complications still arise, see for example Corda et al., 1998). Lam and Kalita (2013) leverage the concept inventories of wordnets to this end, providing an algorithm for dictionary reversal that relies on having wordnet data for one of the two languages involved.

For the creation of new wordnets, a major hurdle is producing the required synset inventory. Many projects follow the *expand model* (Vossen, 1998, p. 83) in which an existing wordnet, usually PWN, is used as a foundation upon which to expand, significantly reducing the required amount of work (Bosch and Griesel, 2017). As a side effect, wordnets that expand from the same wordnet also acquire cross-lingual compatibility through their shared concept inventory. Building on this idea, Bond et al. (2016) introduced the Collaborative InterLingual Index (CILI), an extension of the PWN synset inventory that allows consistent identification of synsets and addition of new synsets and relations to account for concepts and linguistic structures missing from English or Anglocentric cultures. CILI is directly integrated into OMW (Vossen et al., 2016).

3. Creating the Multilingual Sign Language Wordnet

The MSL-WN was started by the EASIER project as a publicly available cross-lingual semantic resource for use in sign language technologies. This section describes work that happened up until the end of the project in December 2023, also documented in project report D6.5 (Bigeard et al., 2024).

During the project, it received three releases: An initial proof-of-concept release providing data for GSL and DGS (Bigeard et al., 2022), a second release covering the remaining project languages (BSL, DSGS, NGT, and LSF) (Bigeard et al., 2023) and a final release introducing the project-external languages PJM and STS (Bigeard et al., 2024).

Inclusion of STS and PJM was made possible through partnerships with the creators of the Swedish Sign Language Dictionary (Svenskt teckenspråkslexikon, 2024) and the Corpus-based Dictionary of Polish Sign Language (Łacheta et al., 2016), who are also co-authoring this article. Similarly, video material and lexical information for the other languages were taken from existing lexical resources, each of which is credited by MSL-WN, including reference links for each individual sign entry in the MSL-WN web interface.

3.1. The annotation interface

To support the annotation efforts for MSL-WN, a custom web interface was developed. It is regularly updated to accommodate new data, add additional features and react to annotator feedback (see Section 4).

The interface provides two annotation perspectives: in *sign view* all meanings (i.e. synsets) of a single sign are annotated, while in *synset view*, one synset is associated with all signs that can represent its meaning. Indices for either perspective provide additional filters, such as listing only synsets that have already been annotated for other sign languages as well, to help annotators focus on what to annotate next. Where frequency information is available from the underlying lexical resource, it is used to sort content, so that more commonly used signs are annotated first. A demonstration of the *sign view* interface is shown in Figure 1. For further information regarding the annotation workflow and interface, see Bigeard et al. (2022) and Bigeard et al. (2024).

3.2. Gloss-based suggestions

To reduce the required amount of manual search for sign-synset connections, the annotation interface provides automatically determined suggestions of likely candidate connections. These suggestions can be generated using different methods. This section discusses the first method, *gloss-based suggestions*[3], introduced in Bigeard et al. (2022), while a new second method, *synonym-based suggestions* is introduced in Section 5.

Due to the lack of established written forms for sign languages, sign language resources commonly supplement the video representation of signs with ID-glosses or keyword translations to textually represent them in lexicon entries, annotations and search interfaces. These are most often produced in the dominant spoken language of the geographic region of the sign language, although some projects also produce additional English versions to facilitate international exchange.

Gloss-based suggestions leverage this information by matching glosses and keywords to lemmas

[3] For brevity, we use the term *gloss-based suggestion* regardless of exactly which spoken language representation is provided for a sign. Depending on the underlying lexical resource, the used text may be a one or several glosses, translational equivalents or other forms of spoken language keyword.

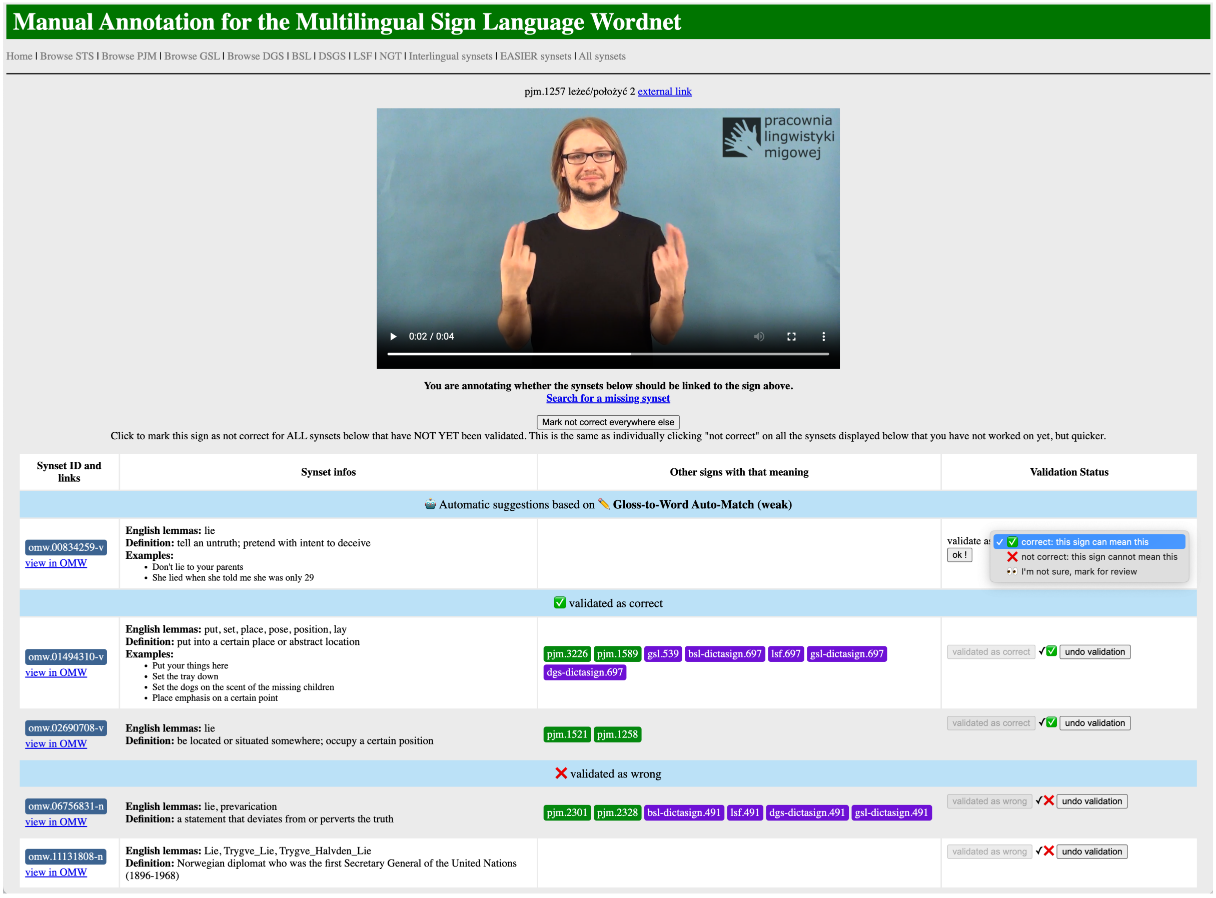

Figure 1: *Sign view* perspective of the MSL-WN annotation interface. The video, ID, glosses and keywords are shown at the top, followed by features for searching missing synsets and mass-rejecting automatic suggestions. The table below lists synsets that are either candidates suggested by automatic systems or have been validated by a human annotator, grouping them accordingly. For each synset, annotators are given links to its entries in the annotation interface and in OMW, lemmas, definitions and examples from available spoken language wordnets, and a list of other signs from the same and other languages already linked to the synset. *Note that the above screenshot has been abridged to allow inclusion in this article.*

found in a wordnet for that spoken language and returning their synsets as candidates. There are several limitations to this approach. As a variant of form-based reverse translation (see Section 2.3), it tends to over-generate for polysemous words, providing many senses that do not apply to the target sign. At the same time, in cases where the lexical data of the sign only provides a single form-level gloss (commonly a best-effort translation for what is assumed the most dominant sense of the sign), secondary senses are still missed out on. Additional technical challenges come from converting glosses to lemmas (Kopf et al., 2022b), handling complex expressions and processing abbreviations and lexicographic addenda.

Nevertheless, given the limited available data and lack of language technologies to support alternative approaches, gloss-based suggestions still represented the best assistive method viable at the time. It was also clearly preferable to requiring annotators to manually look up each synset, a considerably slower approach with its own pitfalls and which in the end also relies on spoken language lemma lookup. Further observations on how gloss-based suggestions affected the annotation process are provided in Section 4.

4. Experiences with Annotation

In this section we share some of the experiences of annotation teams during phase 1 of MSL-WN and how these affected the continuing development of the annotation interface.

4.1. Accommodating different workflows

As was mentioned in Section 3.1, the annotation interface allows annotators to focus on annotating either a specific sign or a specific synset through different annotation views and index lists. Both views were adopted by annotators, with individual annotators showing different preferences for focusing on the *sign view*, *synset view* or dynamically switching between both.

Focusing on a specific view will, of course, have an impact on how the coverage of the dataset progresses, with work through *sign view* contributing to more complete description of all senses of a sign, and *synset* view aiding the interconnectedness of signs for a given concept, although both approaches should largely converge in coverage as annotation progresses. Nevertheless, annotators are encouraged to switch between views, both during exploration of the data and between completed annotation passes.

4.2. Language competence

Language competence of annotators played an important role both regarding the language being annotated and the language that wordnet data is available in.

4.2.1. Annotation language

Ideally, annotators should have L1 competence in the sign language they are annotating. However, given the high demand for L1 signers in linguistic research, some language teams had to fall back on fluent L2 signers as annotators. In these cases, the L2 signers took care of the majority of annotations, consulting the lexical and corpus data of the resource from which MSL-WN took the sign entry, but deferred unclear cases to an L1 signer. The annotation interface was designed to accommodate this need by letting annotators mark entries as needing review, rather than being valid or invalid.

4.2.2. Synset description language

To determine the meaning of a synset, annotators can reference its definition, its list of example sentences and the set of words and signs associated with it. This means that annotators need to be competent in the languages in which this information is provided. While in some cases synset definitions or examples in multiple (spoken) languages are available, often only information in English is available. This can pose challenges for annotators, particularly regarding specific nuances of meaning between closely related synsets, and can add additional requirements regarding the multilingual competence of annotators.

Disambiguating information may also be less complete for some languages. For instance, the GSL team found that Greek synset definitions from BalkaNet were often missing usage examples. This resulted in a degree of uncertainty on the part of annotators as to the accuracy of their choices for sign-synset connections.

At the same time, annotators reported that seeing what signs from other languages were already assigned to a synset could be very helpful when the annotator had competence in that other language as well. This also (slightly) helps to mitigate the issue that definitions are only ever available in spoken languages.

4.2.3. Languages for gloss-based suggestions

Availability of languages also played a big role in how well gloss-based suggestions could be. Most lexical resources provide their text information in the dominant spoken language of the sign language's region. English descriptions may also be provided by some resources, but these are often secondary translations intended to widen access for the international research community.

Where possible, gloss-based suggestions use the regional spoken language. However, as the size of different wordnets can differ strongly and none rival that of PWN, English often has to be used as a fallback option to generate any suggestions at all. For example, only 30% of suggestions for PJM could be generated via Polish keywords and wordnet entries, while 70% were based on English.

Furthermore, verifying the quality of the automated matches between glosses/keywords and wordnet lemmas often fell on annotators as well, as the developers of the matching procedures did not necessarily have the required language competence to do so.

4.3. Changing workflows for GSL

The very first data produced for MSL-WN were annotations of GSL and DGS. During this initial experimental period, different possible workflows were explored in parallel (see Bigeard et al., 2022). As the MSL-WN annotation interface was not yet available, the GSL team used its own internal lexicographic workflows and software (for details, see Vacalopoulou et al., 2022).

Annotation focused on the Greek language part of OMW, a set of 18,000 synsets originally produced for BalkaNet (Tufiş et al., 2004). Using this data allowed the team to produce gloss-based automatic suggestions between Greek wordnet lemmas and the Greek keywords of the GSL lexical database Noema+ (Efthimiou et al., 2016). It also allowed annotators to work with the written language for which they had the strongest language competence, although some issues regarding completeness of wordnet information occurred (see Section 4.2.2). In addition, the known flaws of gloss-based suggestions discussed in Section 3.2 resulted in invalid sign-synset link suggestions in about 30% of the cases.

For the second phase of MSL-WN, the GSL team switched to working with the MSL-WN annotation interface. Adjustment to the new interface was found to be straightforward with no major issues. Having all relevant information gathered in one interface, rather than cross-referencing independent resources, reportedly helped annotators to stay focused on their task. Being shown sense-level synonyms from other (sign) languages (see Figure 1) also helped annotators with knowledge of those languages to distinguish between possible senses (see Section 4.2.2).

4.4. Evolving the interface

The MSL-WN annotation interface and underlying data structures evolved continuously as annotation progressed. For each new language, custom procedures were developed to transfer information from the underlying lexical resource to MSL-WN. The computation of gloss-based suggestions also had to be adjusted to both the language being processed and the structure of lexical resource data (see Section 3.2).

Development of the interface directly took into account annotator feedback. For example, initially, each sign-synset link, including each automatic suggestion, had to be validated individually. Based on a request by the NGT annotation team, a button to summarily reject all automatic suggestions of a specific sign or synset was added (Bigeard et al., 2024). This allows annotators to first validate correct suggestions and then reject the remaining suggestions in one go, speeding up the annotation process. This functionality is currently being extended further to generally allow the dynamic selection of multiple entries for joint submission of a shared validation decision.

4.5. Expanding the sign inventory

The sign inventory of MSL-WN is defined by the lexical sign language resource on which it builds. Accordingly, it inherits any limitations that the lexical resource might have. For instance, the vocabulary of the *Corpus-based Dictionary of Polish Sign Language* (Łacheta et al., 2016) is determined by what signs could be observed in the *Polish Sign Language Corpus* (Kuder et al., 2022). The content of the corpus was in turn affected by its selection of elicitation tasks, so that certain topics are more prevalent than others. As a result, neither corpus nor dictionary cover e.g. slang terms or specialised terminology.

Limitations of vocabulary can directly affect the work of annotators, who might be aware of signs existing for a synset sense, but need to verify through lengthy searches whether it is part of the resource or not.

The best way to address missing vocabulary is, of course, to introduce it. This would either require the production of new lexical materials (which is better achieved by our lexical resource partners than MSL-WN itself) or the inclusion of additional lexical resources. Apart from the general rarity of such resources (see Kopf et al., 2022a), this also introduces the question of how best to identify overlaps in sign inventory between resources of the same language, so as to avoid creating duplicate sign entries.

5. Leveraging Sign-to-Sign Synonymy

Since its inception, the MSL-WN has grown to include over 10,000 verified sign-synset links. To leverage this data for future annotation work, we introduce a new method for generating automatic suggestions of possible sign-synset links. This new method suggests additional meanings based on the sense inventories of other signs that have already been verified to be (partial) synonyms of the sign being annotated. This method can be applied both intra- and cross-lingually. It represents a reversal based on lexical units, which is preferable to the form unit reversal of gloss-based suggestions (see Sections 2.3 and 3.2). It also reduces the dependency on spoken languages as a pivot.

5.1. Implementing synonym suggestions

Synonym-based suggestions are determined as follows: Once two signs are established as having partial synonymy, i.e. they have at least one shared meaning, expressed through both signs being linked to the same synset, there is a reasonable chance that they also share other meanings. When one of the signs is being annotated, the other sign is checked for verified links to additional synsets, which can then be provided as suggestions for the current annotation. A concrete example of this procedure is shown in Figure 2.

In the annotation interface, synonym-based suggestions are grouped separately from gloss-based suggestions and ranked higher (cf. groupings in Figure 1). For sorting the different synonym-based suggestions amongst themselves, the following ranking steps are applied:

1. **Intra- or cross-lingual:** Suggestions are possible both between signs of the same language and across languages, but those from the same language are ranked more highly. Connections that are only established cross-lingually are grouped separately.
2. **Synonym purity:** Apart from annotating valid sign-synset links, annotators can also explicitly mark invalid relations, meaning that a sign

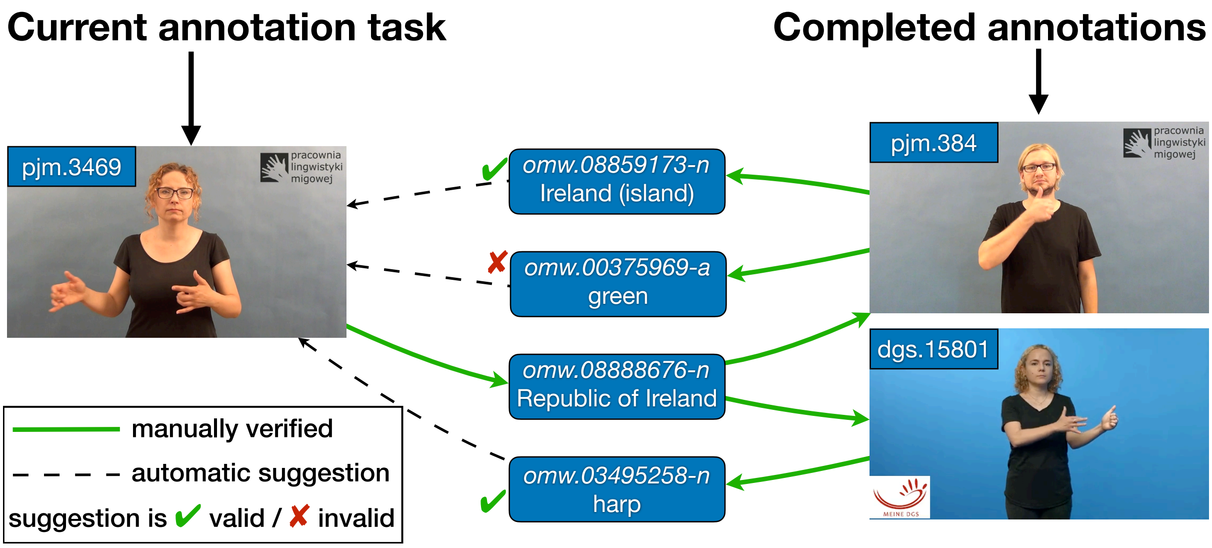

Figure 2: Demonstration of automatic suggestions based on partial synonymy between signs. The sign `pjm3469` is currently being annotated. Once at least one shared meaning is established between signs (here via synset `omw.08888676-n`), the interface suggests other possible shared meanings based on the verified annotations of the other signs. In this case it correctly suggests synsets representing the island of Ireland and a harp, but incorrectly suggests the concept of the colour green. (Image sources: Łacheta et al., 2016; Konrad et al., 2020)

does *not* have the meaning represented by this synset. As a result there can be sign pairs that are confirmed to be only partially synonymous, because while both have valid links to some synsets, there are other synsets that are marked as valid senses of one sign and invalid senses of the other sign. These partial synonyms are ranked lower than signs whose synonym is (so far assumed to be) more complete.

3. **Synonymy strength:** Sign pairs with established synonymy across several senses are more likely to be fully synonymous, so suggestions are ranked higher the more verified synset links a sign pair has in common.

4. **Sign-to-sign quantity:** If a synset is suggested several times through synonym connections with different signs, this candidate is ranked higher than if only one connection suggested it.

These ranking steps are not without flaws, as they can be impacted by how complete the annotations of individual other signs are as well as structural factors of the used sign inventories. For example, phonological variants of a sign are often listed as separate entries, which could inflate their weight in the *sign-to-sign quantity* ranking. Additional information from the underlying lexical resources may be used in future to partially counteract such issues.

5.2. Preliminary annotator feedback

Annotators recently started working with the new synonym-based suggestion feature. Initial observations were that it clearly helps discover senses that were not covered through the spoken language pivot of gloss-based suggestions, in some cases significantly increasing sense coverage of polysemous signs.

It also speeds up annotation of form and dialectal variants, which are usually represented as separate sign entries, but are largely (though not necessarily fully) synonymous. Annotators also welcomed having a more sign-centric approach that helps reduce reliance purely on spoken language data.

In some cases, signs receive a large number of incorrect suggestions when they are linked to a highly polysemous sign. To quickly handle such cases, we are investigating interface improvements to allow a focused inspection of suggestions based on what connection triggered them.

MSL-WN has always kept a record not only of valid sign-synset connections, but also of which connections were verified as incorrect. This is becoming particularly useful in connection with synonym-based suggestions, as it provides hard data on the scope and limits of partial synonymy between signs. This is expected to help linguistic research, for instance in studies comparing how strong the actual synonymy between signs from different languages is.

Published	GSL	PJM	STS	Total
Signs	2,949	2,415	2,706	14,028
Synsets	5,760	3,626	1,981	16,534
Links	7,499	6,486	2,810	24,367

Table 1: Counts of signs, synsets and sign-synset links included in the public MSL-WN dataset. Shows the languages currently being worked on and the sums for all languages.

6. New Dataset Release

While the EASIER project has ended, work on MSL-WN continues. The first step is the annotation of additional data for PJM, STS and GSL. The current state of this on-going effort has already been published to the MSL-WN dataset.[4] Statistics for recently updated languages can be seen in Table 1.

This new round of annotations is produced in concert with the interface and workflow improvements presented in this article. Supported by the new sign-to-sign synonym suggestion feature, annotators have been encouraged to focus on synsets that have already been verified for other signs. Compared to the release described in Bigeard et al. (2024), the number of synsets covered by at least two languages rose by 79% to 3,361 and those covered by at least three or four languages rose by 59% (1,472) and 51% (617), respectively.

7. Conclusion

This article presents the current progress of the Multilingual Sign Language Wordnet (MSL-WN). It describes the experiences gathered so far and how workflows and software support were adjusted according to annotator feedback. A major change is the recent addition of a new automatic suggestion feature that leverages established partial synonymy between signs. This helps reduce the dependence on spoken language form-level suggestions.

Annotation of Multilingual Sign Language Wordnet (MSL-WN) is on-going. A new dataset release which introduces new annotations for GSL, PJM, and STS accompanies this article. Work on these annotations is also used to gather initial feedback on the impact of the new interface features.

In future, a varied number of improvements are planned. Some of these regard the fluency of the annotation workflow, such as allowing more flexible simultaneous verification of sets of suggestions. Others will address questions of how to integrate multiple lexical resources for one sign language and how to represent sign language phenomena such as classifiers and incorporation in wordnet.

[4] https://doi.org/10.25592/uhhfdm.14190

8. Acknowledgements

We would like to thank Neil Fox, Kearsy Cormier, Onno Crasborn, Lianne Westenberg, Sarah Ebling and Laure Wawrinka for their work in prior phases of MSL-WN on which this article builds. We would like to thank Gabriele Langer for valuable discussions regarding lexicographic practices.

This publication has been produced in part in the context of the joint research funding of the German Federal Government and Federal States in the Academies' Programme, with funding from the Federal Ministry of Education and Research and the Free and Hanseatic City of Hamburg. The Academies' Programme is coordinated by the Union of the Academies of Sciences and Humanities.

This work is supported in part by the EASIER (Intelligent Automatic Sign Language Translation) Project. EASIER has received funding from the European Union's Horizon 2020 research and innovation programme, grant agreement n° 101016982.

This work is supported in part by INRIA within the project Defi COLaF.

AK is funded by the DFG Priority Programme 2392 Visual Communication (ViCom, 2022-2025), Frankfurt am Main, Germany.

9. Ethical considerations

Deaf signers were consulted in the development of MSL-WN and its annotation interface. Annotation for MSL-WN was performed either by deaf signers or in consultation with them.

MSL-WN is a resource for sign linguistic research. Its license excludes commercial uses. To work towards ethical applications of sign language technologies, its dataset download page provides an advisory for users to observe the best practice recommendations by Fox et al. (2023).

10. Limitations

The MSL-WN is a work in progress. Individual sign entries are not guaranteed to yet have synset links for all their senses. Correspondingly, synset entries should not be expected to contain a complete lists of matching synonyms for any given language.

In addition to items verified by human annotator, MSL-WN also contains automatically detected sign-synset links for signs that are presumed to be monosemous. The dataset explicitly marks whether an entry was human- or machine-verified.

As the MSL-WN builds on the lexical inventory of several other resources, the editorial decisions of those resources can affect how signs are represented in MSL-WN, e.g. regarding how fine-grained the distinction of lexical variants is.

11. Bibliographical References

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

Sam Bigeard, 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, EASIER Consortium, Hamburg, Germany.

Sam Bigeard, 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, Laure Wawrinka, Johanna Mesch, Thomas Björkstrand, Anna Kuder, and Joanna Wójcicka. 2024. Extended interlingual index for the project's core sign languages and languages covered in WP9. Project Note D6.5, EASIER Consortium, Hamburg, Germany.

Francis Bond and Kyonghee Paik. 2012. A survey of WordNets and their licenses. In *Proceedings of the 6th Global WordNet Conference*, pages 64–71, Matsue, Japan.

Francis Bond, 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.

Sonja E. Bosch and Marissa Griesel. 2017. Strategies for building wordnets for under-resourced languages: The case of African languages. *Literator*, 38(1):1–12.

Alessandra Corda, Vincenzo Lo Cascio, and Massimiliano Pipolo. 1998. Automatic reversal of a bilingual dictionary: Implications for lexicographic work. In *Proceedings of the 8th EURALEX International Congress*, pages 433–443.

Thierry Declerck and Sussi Olsen. 2023. Linked open data compliant representation of the interlinking of Nordic wordnets and sign language data. In *Proceedings of the Second Workshop on Resources and Representations for Under-Resourced Languages and Domains (RESOURCEFUL-2023)*, pages 62–69, Tórshavn, the Faroe Islands. Association for Computational Linguistics.

Dicta-Sign Consortium. 2012. Sign-Wiki prototype. Project deliverable D6.7, Dicta-Sign Consortium.

Sarah Ebling, Katja Tissi, and Martin Volk. 2012. Semi-automatic annotation of semantic relations in a Swiss German Sign Language lexicon. In *Proceedings of the LREC2012 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon*, pages 31–36, Istanbul, Turkey. European Language Resources Association (ELRA).

Eleni Efthimiou, 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 *Proceedings of the LREC2016 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining*, pages 63–68, Portorož, Slovenia. European Language Resources Association (ELRA).

Eleni Efthimiou, 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 *Proceedings of the LREC2012 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon*, pages 37–44, Istanbul, Turkey. European Language Resources Association (ELRA).

Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. The MIT Press.

Neil Fox, Bencie Woll, and Kearsy Cormier. 2023. Best practices for sign language technology research. *Universal Access in the Information Society*.

Birgit Hamp 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*, pages 9–15, Madrid, Spain. Association for Computational Linguistics.

Reiner Konrad, 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, Oliver Böse, Elena Jahn, and Marc Schulder. 2020. *MEINE DGS – annotiert. Öffentliches Korpus der Deutschen Gebärdensprache, 3. Release / MY DGS – annotated. Public Corpus of German Sign Language, 3rd release*. Universität Hamburg. PID https://doi.org/10.25592/dgs.corpus-3.0.

Maria Kopf, Marc Schulder, and Thomas Hanke. 2022a. The Sign Language Dataset Compendium: Creating an overview of digital linguistic resources. In *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources*, pages 102–109, Marseille, France. European Language Resources Association (ELRA).

Maria Kopf, Marc Schulder, Thomas Hanke, and Sam Bigeard. 2022b. Specification for the harmonization of sign language annotations. Project Note D6.2, EASIER Consortium, Hamburg, Germany.

Anna Kuder, 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*, pages 118–123, Marseille, France. European Language Resources Association (ELRA).

Joanna Łacheta, 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.

Khang Nhut Lam and Jugal Kalita. 2013. Creating reverse bilingual dictionaries. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 524–528, Atlanta, Georgia. Association for Computational Linguistics.

Gabriele Langer and Marc Schulder. 2020. Collocations in sign language lexicography: Towards semantic abstractions for word sense discrimination. In *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*, pages 127–134, Marseille, France. European Language Resources Association (ELRA).

Jurie Le Roux, Koliswa Moropa, Sonja Bosch, and Christiane Fellbaum. 2008. Introducing the African languages wordnet. In *Proceedings of The Fourth Global WordNet Conference*, pages 269–280, Szeged, Hungary. University of Szeged, Department of Informatics.

Colin P. Lualdi, 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*, pages 63–72, Potchefstroom, South Africa. South African Centre for Digital Language Resources (SADiLaR).

Willy Martin. 2013. Reversal of bilingual dictionaries. In Rufus H. Gouws, Ulrich Heid, Wolfgang Schweickard, and Herbert E. Wiegand, editors, *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*, volume 5/4 of *Handbooks of Linguistics and Communication Science*, pages 1445–1455. De Gruyter Mouton, Berlin.

Silke Matthes, Thomas Hanke, Anja Regen, Jakob Storz, Satu Worseck, Eleni Efthimiou, Athanasia-Lida Dimou, Annelies Braffort, John Glauert, and Eva Safar. 2012. Dicta-Sign – building a multilingual sign language corpus. In *Proceedings of the LREC2012 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon*, pages 117–122, Istanbul, Turkey. European Language Resources Association (ELRA).

George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller. 1990. Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, 3(4):235–244.

Bolette S. Pedersen, 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. *Language Resources and Evaluation*, 43:269–299.

Siegmund Prillwitz, Thomas Hanke, Susanne König, Reiner Konrad, Gabriele Langer, and Arvid Schwarz. 2008. DGS Corpus project – development of a corpus based electronic dictionary German Sign Language / German. In *Proceedings of the LREC2008 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign*

Language Corpora, pages 159–164, Marrakech, Morocco. European Language Resources Association (ELRA).

Umar Shoaib, Nadeem Ahmad, Paolo Prinetto, and Gabriele Tiotto. 2014. Integrating MultiWordNet with Italian Sign Language lexical resources. *Expert Systems with Applications*, 41(5):2300–2308.

Svenskt teckenspråkslexikon. 2024. *Swedish Sign Language Dictionary online*. Department of Linguistics, Stockholm University.

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

Thomas Troelsgård and Jette Hedegaard Kristoffersen. 2018b. Building a sign language corpus – problems and challenges: The Danish Sign Language Corpus and Dictionary. Abstract published at the XVIII EURALEX International Congress Lexicography in Global Contexts.

Dan Tufiş, Dan Cristea, and Sofia Stamou. 2004. BalkaNet: Aims, methods, results and perspectives. A general overview. *Romanian Journal of Information Science and Technology*, 7(1–2):9–43.

Anna Vacalopoulou, Eleni Efthimiou, Stavroula-Evita Fotinea, Theodoros Goulas, Athanasia-Lida Dimou, and Kiki Vasilaki. 2022. Organizing a bilingual lexicographic database with the use of WordNet. In *Dictionaries and Society. Proceedings of the XX EURALEX International Congress, 12-16 July 2022, Mannheim, Germany*, pages 357–366. IDS-Verlag.

Piek Vossen, editor. 1998. *EuroWordNet: A multilingual database with lexical semantic networks*. Springer Netherlands, Dordrecht.

Piek Vossen, Francis Bond, and John McCrae. 2016. Toward a truly multilingual GlobalWordnet grid. In *Proceedings of the 8th Global WordNet Conference (GWC)*, pages 424–431, Bucharest, Romania. Global Wordnet Association.