







Introducing Sign Languages to a Multilingual Wordnet: Bootstrapping Corpora and Lexical Resources of Greek Sign Language and German Sign Language

Sam Bigeard¹, Marc Schulder¹, Maria Kopf¹, Thomas Hanke¹, Kyriaki Vasilaki², Anna Vacalopoulou², Theodoros Goulas², Athanasia-Lida Dimou², Stavroula-Evita Fotinea², Eleni Efthimiou² ¹ University of Hamburg, Institute of German Sign Language and Communication of the Deaf, Germany ² Institute for Language and Speech Processing (ILSP), Athena Research Center, Greece

Beta version available at sign-net.meine-dgs.de





Greek:

τράπεζα

Bulgarian:

банка



synset

channels money



French: synset banque English: sloping land bank (especially beside a body of water)

Wordnets we use:

- · Open Multilingual Wordnet: 34 connected wordnets in different languages
- Germanet: 28,564 synsets mapped to OMW
- · Greek Wordnet: 18,049 synsets

Sign resources we use:

- DGS Corpus
- Polytropon

Why?

Goal: Connect existing sign language lexicons & make them machine-readable

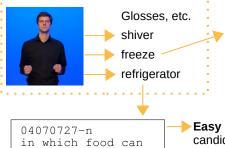
Why a Wordnet?

- Precise senses, accessible through definition, words and relationships
- Sense is identified by number, not translation to spoken language
- Machine-readable. Libraries to manipulate it already exist

Why expand on existing wordnets rather than create one from scratch?

- Easier and quicker to build
- More accessible to future users
- Compatible with many resources: languages, pictures, etc.

How?



- 00374135-v change to ice
- 01834730-v stop moving
- 00012613-v suddenly behave coldly and formally

Easy case: Only one candidate synset. Low risk of mistranslation: Automatic validation possible

Many candidate synsets: Manual validation. The process varies depending on what kind of data the resource offers.

GSL dictionary:

Example sentence for select meaning(s). Expert required for other potential meanings.

DGS corpus:

Occurrences of sign in various contexts. Provides potential examples for more senses, but intended sense not necessarily explicit.

Synsets linked in one language are then prioritised in the other to improve interconnection.

Current Progress

be stored at low

temperatures

		GSL validated	DGS candidates	DGS validated	GSL/DGS overlap
	distinct synsets	4214	27,020	969	278
	distinct signs	1819	11,856	2230	N/A
	sign-synset pairs	4347	138,518	2230	N/A

ready to be used in applications validated. candidates: automatic matches, needs to be verified synsets linked to both GSL and DGS overlap:

Why are the ratios different?

Corpus-based (DGS): Many signs with overlapping meanings in corpus. One synset associated with a sign propagates to synonymous signs, resulting in more candidate signs.

Dictionary-based (GSL): Dictionary created with focus on providing many concepts with at least one sign. Results in faster validation of many synsets, but fewer signs per synset.

Work in progress: We expect the two datasets to evolve toward many sign-synset pairs, as is seen currently in the DGS candidates column.

ter presented at the 10th Workshop on the resentation and Processing of Sign guages: Multilingual Sign Language ources. LREC 2022. Marseille, France. e 25th, 2022 Funded by the Horizon 2020 Framework Programme of the European Union