

The Multilingual Sign Language Wordnet

Sam Bigeard - 7/5/24

INRIA Nancy, France

Presenting work done in IDGS, University of Hamburg

Also presenting the work of Marc Schulder, Maria Kopf, Thomas Hanke,
and many annotators

Sam Bigeard

Post-doc at INRIA Nancy, France

- Bachelor in linguistics
- Master and PhD in NLP
- Thesis and post-doc in medical data
- Post-doc in SL at IDGS Hamburg for 3 years
- Post-doc in resources for under-resources languages at INRIA Nancy

- Why we started this work
- What is Wordnet?
- How can it help?
- The multilingual SL Wordnet
- How it was built

Intelligent
Automatic
Sign
Language
Translation



EASIER Project

Ran for 3 years 2020-2023

Horizon 2020 project

Goal : multilingual translation between

NGT, DGS, DSGS, GSL, BSL, LSF, LIS

and surrounding spoken languages

any pair, any direction

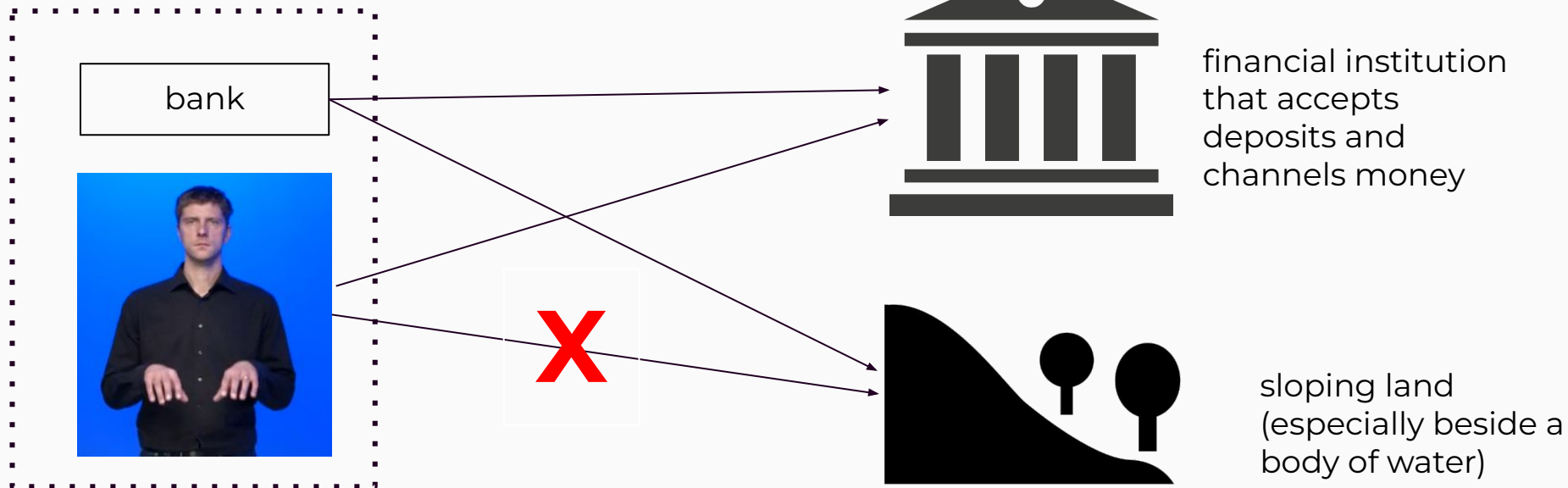
Why we started this work



Why we started this work

Our problem : senses need to be searchable across languages and disambiguated

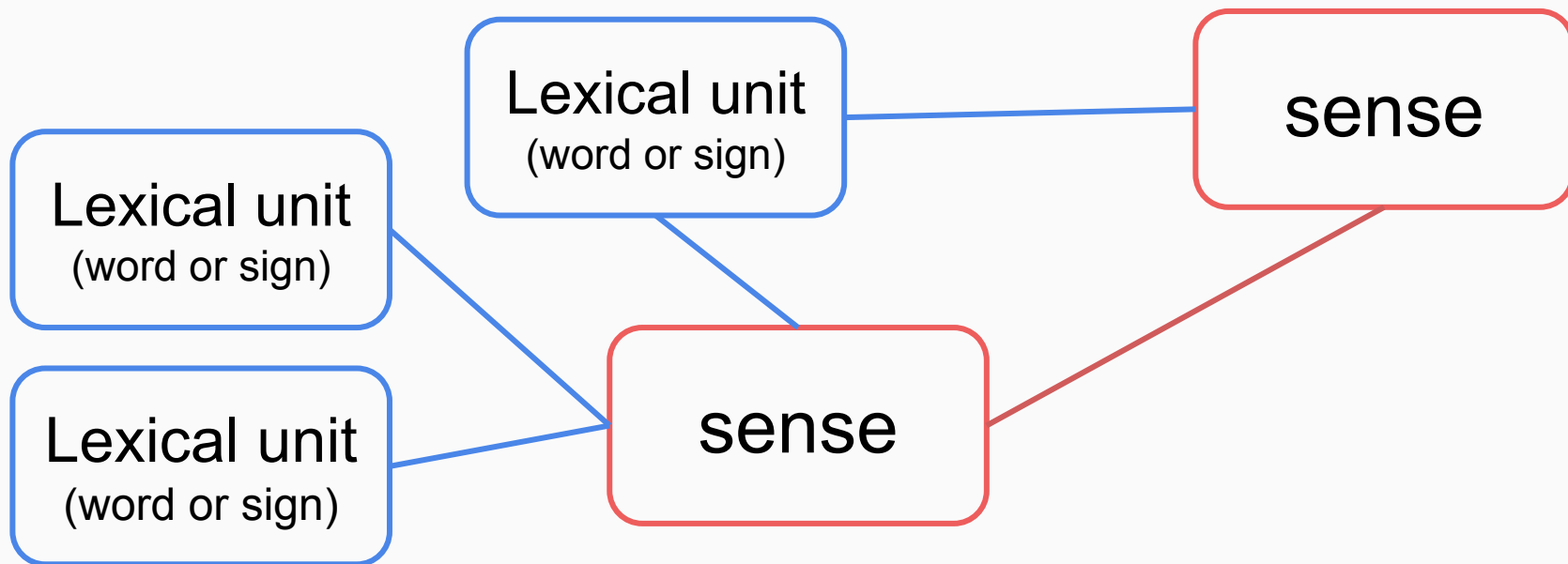
SL lexicon

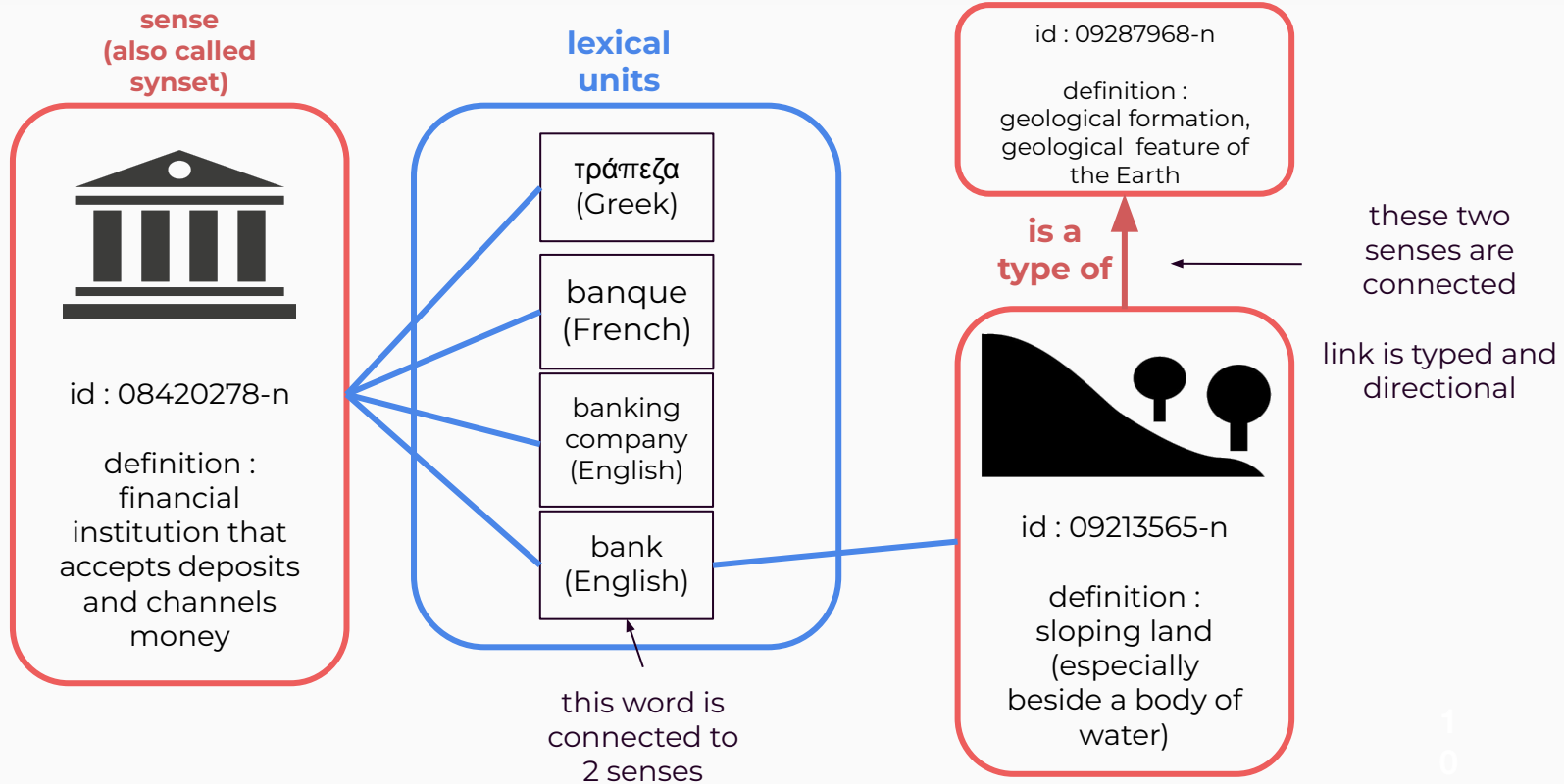


How to represent the sense in a Signbank?

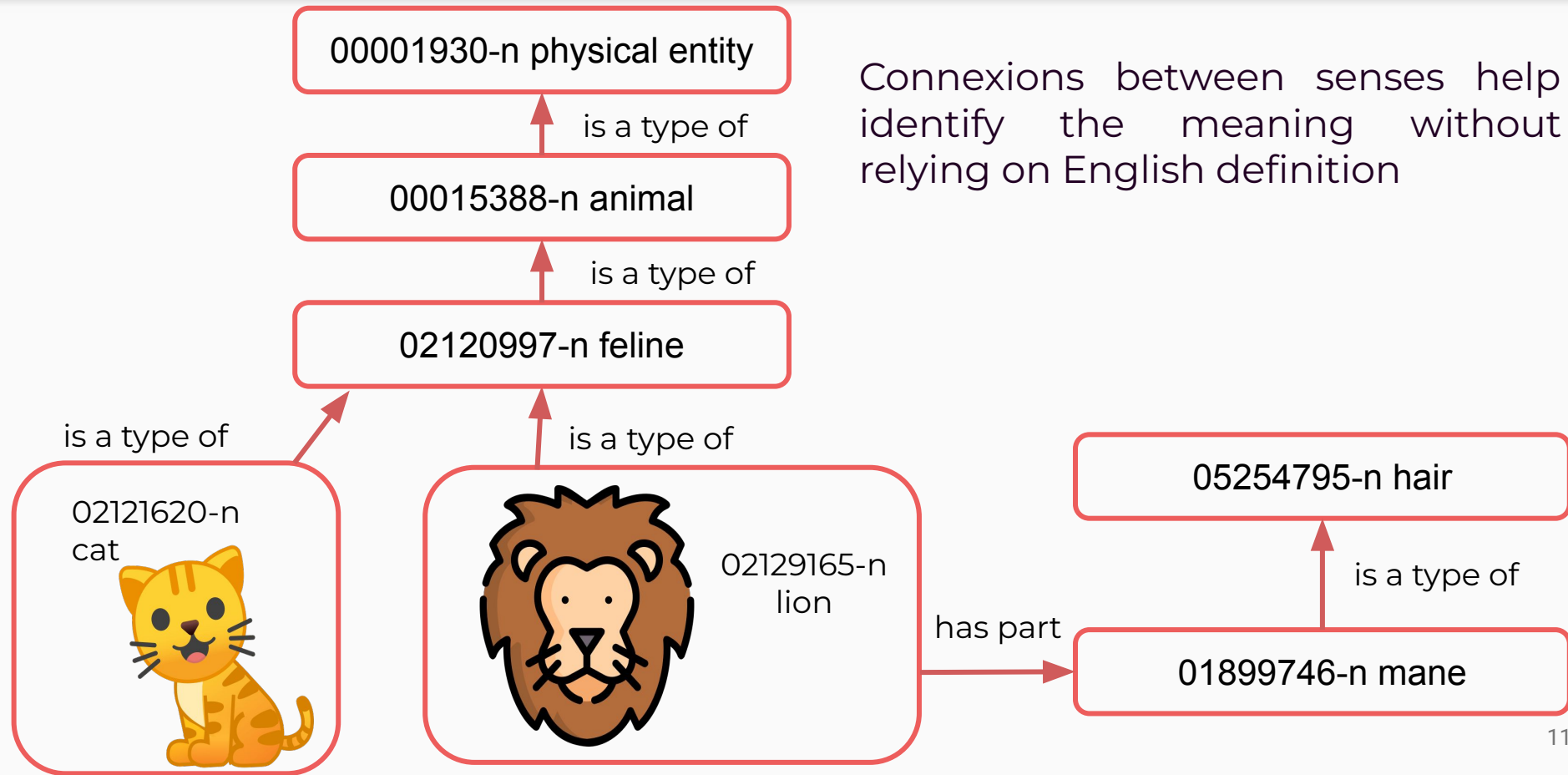
- Bilingual strategy : Spoken language equivalents
 - but: not precise, subject to mistranslation, depends on contact language
- Monolingual strategy : Definitions and examples
 - but: extra work
- Wordnet strategy : Semantic links

Wordnet = **senses** + **lexical units** + links

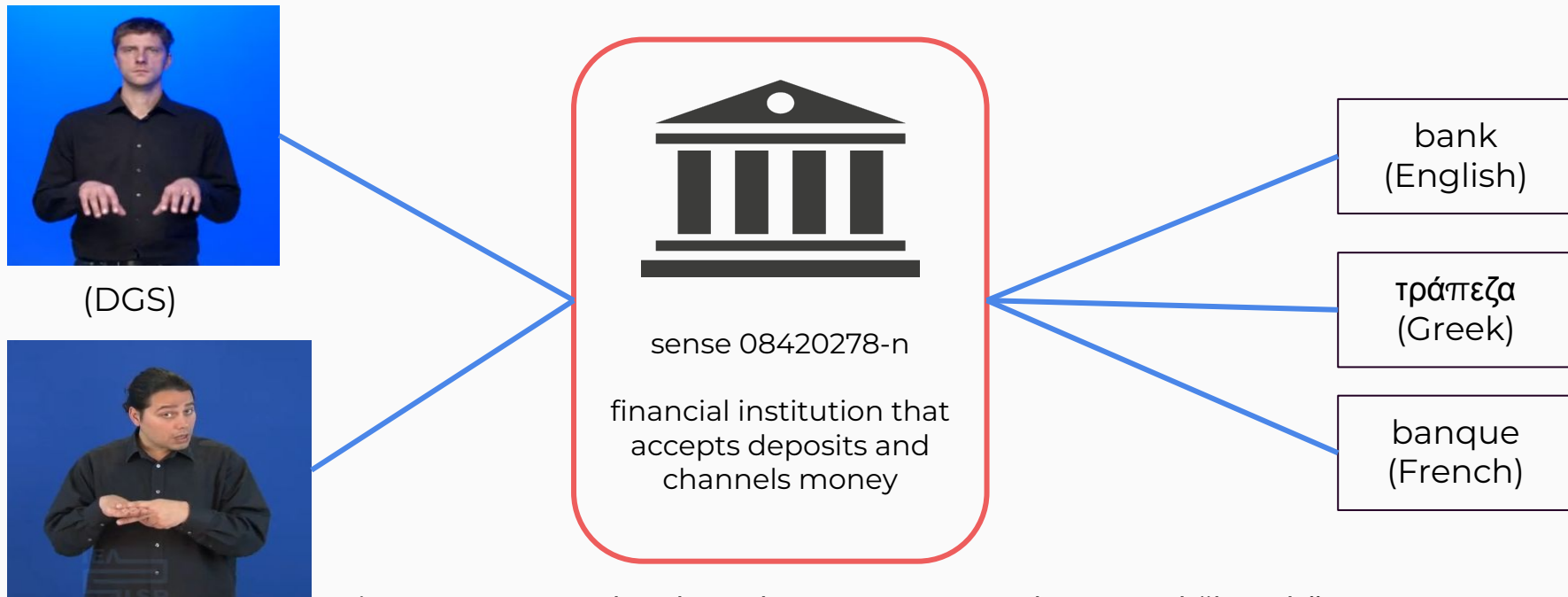




Connexions between senses help identify the meaning without relying on English definition



We simply add new **lexical units** that contain a sign instead of a word



Signs are attached to the sense, not the word “bank”
-> avoids ambiguity of translation



sense 08420278-**n**

financial institution that
accepts deposits and
channels money

This is a part of speech
Problem for SL

- We can associate a sign with several synsets to cover several POS
- Wordnet people are open to make POS-agnostic synsets

Scope of Wordnet:

117 000 synsets

all have definition and lemmas in English

most have example sentences in English

Lemmas in 28 languages, varying in coverage between languages

Linked to many other resources for more content, like pictures

Princeton's Wordnet, [Miller 1995, Fellbaum 1998](#)

Open Multilingual Wordnet, [Bond and Foster 2013](#)

let's play!

<https://compling.upol.cz/ntumc/cgi-bin/wn-gridx.cgi?gridmode=grid>

and image demo



Summary, Wordnet can give you :

- Large sense inventory
 - No need to create senses or write definitions
 - Spoken language lemmas and synonyms, useful for search
 - Multilingual
 - Can help identify missing senses
- Semantic links between the senses : help describe it, make thematic groups
- Pictures

The Multilingual Sign Languages Wordnet

21,635 links between 13,602 signs and 15,143 synsets in 8 languages
still growing

- Sign Language of the Netherlands
- German Sign Language
- Greek Sign Language
- British Sign Language
- French Sign Language
- Swiss German Sign Language
- Swedish Sign Language
- Polish Sign Language

DGS



dgs.9761 AUTUMN1A^

en

fall

en

autumn

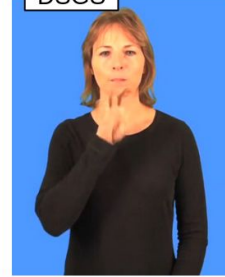
de

Herbst

omw.15236859-n



DSGS



dsgs.1353 HERBST_1A

NGT



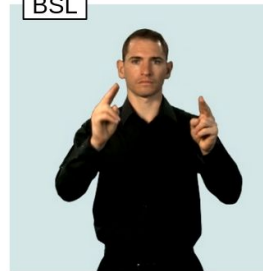
ngt.438 HERFST

GSL



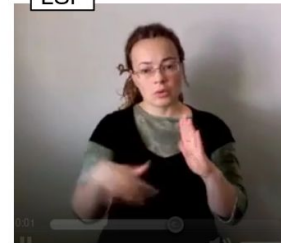
gsl.6902 φθινόπωρο

BSL



bsl.2695 AUTUMN

LSF



lsf.56 AUTOMNE

omw.00049433-r now | in the historical present; at this point in the narration of a series of past events

- [bsl.4114](#) current, now, currently, nowadays, today, present
- [dgs.80580](#) HIER-JETZT1^
- [dgs.14501](#) JETZT2^
- [dgs.15532](#) JETZT3^
- [gsl.2951](#) σήμερα

omw.00051848-r directly, straight, direct | without deviation

- [dgs.14265](#) RICHTUNG1^
- [dgs.14543](#) ZIEL2^
- [dgs.79772](#) ZIEL4^
- [lsf.856](#) TOUT DROIT

omw.00056930-v give birth, deliver, bear, birth, have | cause to be born

- [dgs.13995](#) GEBURT1A^
- [gsl.686](#) γεννώ
- [lsf.374](#) DONNER NAISSANCE
- [ngt.1212](#) GEBOREN-A / BORN-A

omw.00058519-n exit | the act of going out

- [dgs.13504](#) LASSEN2A^
- [dgs.40697](#) LASSEN2C^
- [dgs.26982](#) SPRINGEN2^
- [gsl.1159](#) έξοδος

dgs.3603 RUND3A^

View more data about this sign in its original resource: [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	<ul style="list-style-type: none"> round trip 	a trip to some place and back again		Automatic validation	
omw.07873807-n omw link internal link	<ul style="list-style-type: none"> 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	
omw.06793231-n omw link internal link	<ul style="list-style-type: none"> sign 	a public display of a message	<ul style="list-style-type: none"> he posted signs in all the shop windows 	Automatic validation	GSL

Data available in CSV and NLTK format

```
wn.synset_from_pos_and_offset("n",2129165).lemma_names("ngt")
```

```
>>> ['https://signbank.cls.ru.nl/dictionary/protected_media/  
glossvideo/NGT/LE/LEEUEW-B-22.mp4',  
      'https://signbank.cls.ru.nl/dictionary/protected_media/  
glossvideo/NGT/LE/LEEUEW-A-1759.mp4']
```

```
>>> ['LEEUEW-B', 'LION-B', 'LEEUEW-A', 'LION-A']
```

```
sense=wn.synsets("LION-B",lang="ngt")[0]
```

```
sense.offset()
```

```
>>> 2129165
```

```
sense.definition()
```

```
>>> large gregarious predatory feline of Africa and India having a  
tawny coat with a shaggy mane in the male
```

let's play!

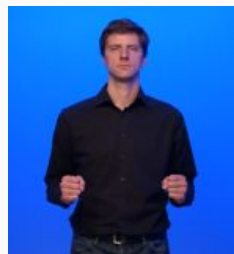
<http://sign-net.meine-dqs.de>



How it was done

Making the SL Wordnet

SL lexicon



glosses, keywords, etc
in spoken language

- refrigerator
- freeze
- shiver

step 1: automatically
find candidate synsets

- 04070727-n in which food
can be stored at low
temperatures

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



step 2: if only one
quality sense, auto
validation

step 3: manual validation

Current annotation task



— manually verified
- - - - automatic suggestion
suggestion is ✓ valid / ✗ invalid

✓ *omw.08859173-n*
Ireland (island)

✗ *omw.00375969-a*
green

omw.08888676-n
Republic of Ireland

✓ *omw.03495258-n*
harp

Completed annotations



work of Marc Schuder

With the generous work of :

Conception and DGS data: Sam Bigeard, Marc Schulder, Maria Kopf and Thomas Hanke (*Institute of German Sign Language and Communication of the Deaf (IDGS), University of Hamburg, Germany*)

GSL data: Kiriaki Vasilaki, Anna Vacalopoulou, Theodor Goulas, Athanasia–Lida Dimou, Stavroula–Evita Fotinea and Eleni Efthimiou (*Institute for Language and Speech Processing (ILSP), Athena Research Center, Greece*)

BSL data: Neil Fox and Kearsy Cormier (*Deafness, Cognition and Language Research Centre (DCAL), University College London, United Kingdom*)

NGT data: Onno Crasborn and Lianne Westenberg (*Centre for Language Studies, Radboud University, Netherlands*)

DSGS data: Sarah Ebling and Laure Wawrinka (*Department of Computational Linguistics, University of Zurich, Switzerland*)

STS data: Johanna Mesch and Thomas Björkstrand (*Department of Linguistics, Stockholm University, Sweden*)

PJM data: Anna Kuder (*Department of Linguistics, University of Cologne, Germany*) and Joanna Wójcicka (*Institute of Applied Linguistics, Warsaw University, Poland*)

Learn more at the tutorial, 2pm tomorrow !

Thank you !