

D6.1 OVERVIEW OF DATASETS FOR THE SIGN LANGUAGES OF EUROPE

Revision: v1.0

Work Package	WP6
Task	T6.1
Due date	30/06/2021
Submission date	30/07/2021
Deliverable lead	Universität Hamburg
Version	1.0
DOI (latest version)	10.25592/uhhfdm.9560
Authors	Maria Kopf, Marc Schulder, Thomas Hanke (Universität Hamburg – UHH)
Reviewers	Eleni Efthimiou, Evita Fotinea (Athena Research Center – ATHENA), Mélanie Hénault-Tessier (Interpretis – INT)

Abstract	<p>This document identifies linguistic corpora that can be explored as high-quality training data for automatic translation within EASIER (as opposed to loosely aligned broadcast data). For each data set, the document lists what parts of the data are available under what access conditions. It also lists the elicitation formats used in several corpora in order to identify those parts of the available corpora that could be explored to build multilingual resources.</p> <p>In order to support the construction of an interlingual index across European sign languages, the document also lists lexical resources (lexical databases and dictionaries) available and their characteristics.</p>
Keywords	linguistic sign language corpora, lexical databases for sign languages, sign language dictionaries, elicitation formats for sign language data collections, interlingual index



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	09/06/2021	Data presented to project partners in internal workshop	Maria Kopf, Marc Schulder, Thomas Hanke (UHH)
V0.2	23/07/2021	Document draft integrating feedback from partners and external data creators	Maria Kopf, Marc Schulder, Thomas Hanke (UHH)
V0.3	24/07/2021	Internal Review 1	Eleni Efthimiou, Evita Fotinea (ATHENA)
V0.4	28/07/2021	Internal Review 2	Mélanie Hénault-Tessier (INT)
V1.0	29/07/2021	Camera-ready submission DOI: 10.25592/uhhfdm.9561	Maria Kopf, Marc Schulder, Thomas Hanke (UHH)

DISCLAIMER

The information, documentation and figures available in this deliverable are written by the “Intelligent Automatic Sign Language Translation” (EASIER) project’s consortium under EC grant agreement 101016982 and do not necessarily reflect the views of the European Commission.

The European Commission is not liable for any use that may be made of the information contained herein.

COPYRIGHT NOTICE

© 2021 EASIER Consortium

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

- * R: Document, report (excluding the periodic and final reports)
- DEM: Demonstrator, pilot, prototype, plan designs
- DEC: Websites, patents filing, press & media actions, videos, etc.
- OTHER: Software, technical diagram, etc

EXECUTIVE SUMMARY

This document lists linguistic resources that can be used within EASIER and the extent to which they are publicly available. More specifically, it lists:

- linguistic corpora of European sign languages of substantial size that can be used as high-quality training data for automatic translation
- data collection tasks used in more than one of these linguistic corpora
- lexical resources of European sign languages in order to support the construction of an interlingual index (T6.2).

The document has been compiled in such a way that it can be updated regularly in order to include additional resources made available and progress on resources listed as well as to correct omissions or errors in the data listed through feedback from data owners.

CONTENTS

Executive Summary	3
List of Figures	6
Abbreviations	10
1 Introduction	11
2 Methodology	13
2.1 Collection of Metadata	13
2.2 Selection Criteria	13
2.2.1 Corpora	14
2.2.2 Data Collection Tasks	14
2.2.3 Lexical Resources	14
2.2.4 Additional Information	14
3 Resources: Corpora	15
3.1 British Sign Language Corpus	15
3.2 CORLSE	16
3.3 Corpus LSFB	17
3.4 Corpus NGT	18
3.5 Corpus of Finnish Sign Language	19
3.6 Corpus Vlaamse Gebarentaal	20
3.7 CREAGEST	21
3.8 Danish Sign Language Corpus	22
3.9 DGS Corpus	23
3.10 Dicta-Sign Corpus	24
3.11 Dicta-Sign-GSL-v2	25
3.12 Dicta-Sign-LSF-v2	26
3.13 ECHO Corpus	27
3.14 Giving Cognition a Hand Corpus	28
3.15 Hungarian Sign Language Corpus	28
3.16 IPROSLA Corpus	29
3.17 Italian Sign Language Corpus	30
3.18 MEDI-API-SKEL	31
3.19 PJM Corpus	32
3.20 POLYTROPON	33
3.21 SIGNOR Corpus	34
3.22 Signs of Ireland	35
3.23 Spanish Sign Language Corpus	36
3.24 Swedish Sign Language corpus	36
3.25 VID I Sign Space Corpus	37
3.26 Visibase Corpus	38
4 Resources: Data Collection Tasks	39
4.1 Silvester and Tweety	39

4.2	Frog Story	40
4.3	Pear Story	42
4.4	Signs Movie	43
4.5	Charlie Chaplin	44
4.6	Fire Alarm Story	45
4.7	Snowman Story	46
4.8	Mr. Bean	46
4.9	Retelling of fables	47
4.10	Present yourself	48
4.11	Sign Name	49
4.12	Jokes	50
4.13	Free conversation	51
4.14	Deaf life experiences	53
4.15	Your region	54
4.16	Subject areas	55
4.17	What did you do when it happened	58
4.18	Role play	59
4.19	Volterra picture task	59
4.20	Route description	60
4.21	Calendar	62
4.22	Warning and prohibition signs	63
4.23	Describe process	64
4.24	Lexical elicitation	65
4.25	Diachronic changes	66
4.26	Language awareness	67
5	Resources: Lexical Resources	69
5.1	ASL Signbank	69
5.2	ASL-LEX	70
5.3	BSL SignBank	71
5.4	CDPSL	71
5.5	CZJ domain specific Lexicon	72
5.6	Danish Sign Language Dictionary	72
5.7	DGS Corpus types list	73
5.8	Dicta-Sign Lexicon	74
5.9	Dictio	75
5.10	DILSE	75
5.11	DIZLIS	76
5.12	DW-DGS	76
5.13	e-LIS	77
5.14	Elix	77
5.15	Finland-Swedish SignWiki	78
5.16	Finnish Signbank	78
5.17	Finnish SignWiki	79
5.18	GaLex	80
5.19	GLex	80
5.20	Global Signbank - LSFB	81
5.21	Global Signbank - NGT	81
5.22	Global Signbank - NTS	82

5.23	Global Signbank - VGT	82
5.24	HLex	83
5.25	LedaSila	83
5.26	Lex-LSFB	84
5.27	LSE-Sign	84
5.28	LSFB en ligne	85
5.29	NOEMA	86
5.30	Ocelles	86
5.31	PLex	87
5.32	POLYTROPON	87
5.33	sematos	88
5.34	SGB-FSS Lexicon	88
5.35	SignStudy	89
5.36	SLex	89
5.37	Slovar SZJ	90
5.38	SSLD	90
5.39	SUVI or Basic Dictionary of FinSL	91
5.40	Swiss German Sign Language Lexicon	92
5.41	TLex	92
6	Conclusions	94
	Acknowledgements	95
	References	96

LIST OF TABLES

3.1	Fact Sheet: British Sign Language Corpus	15
3.2	Fact Sheet: CORLSE	16
3.3	Fact Sheet: Corpus LSFB	17
3.4	Fact Sheet: Corpus NGT	18
3.5	Fact Sheet: Corpus of Finnish Sign Language	20
3.6	Fact Sheet: Corpus Vlaamse Gebarentaal	21
3.7	Fact Sheet: CREAGEST	22
3.8	Fact Sheet: Danish Sign Language Corpus	23
3.9	Fact Sheet: DGS Corpus	24
3.10	Fact Sheet: Dicta-Sign Corpus	25
3.11	Fact Sheet: Dicta-Sign-GSL-v2	26
3.12	Fact Sheet: Dicta-Sign-LSF-v2	26
3.13	Fact Sheet: ECHO Corpus	27
3.14	Fact Sheet: Giving Cognition a Hand Corpus	28
3.15	Fact Sheet: Hungarian Sign Language Corpus	29
3.16	Fact Sheet: IPROSLA Corpus	30
3.17	Fact Sheet: Italian Sign Language Corpus	30
3.18	Fact Sheet: MEDI-API-SKEL	31
3.19	Fact Sheet: PJM Corpus	32
3.20	Fact Sheet: POLYTROPON	33
3.21	Fact Sheet: SIGNOR Corpus	34
3.22	Fact Sheet: Signs of Ireland	35
3.23	Fact Sheet: Spanish Sign Language Corpus	36
3.24	Fact Sheet: Swedish Sign Language corpus	36
3.25	Fact Sheet: VIDI Sign Space Corpus	37
3.26	Fact Sheet: Visibase Corpus	38
4.1	Fact Sheet: Silvester and Tweety	39
4.2	Corpus uses of task “Silvester and Tweety”	39
4.3	Fact Sheet: Frog Story	40
4.4	Corpus uses of task “Frog Story”	41
4.5	Fact Sheet: Pear Story	42
4.6	Corpus uses of task “Pear Story”	43
4.7	Fact Sheet: Signs Movie	43
4.8	Corpus uses of task “Signs Movie”	44
4.9	Fact Sheet: Charlie Chaplin	44
4.10	Corpus uses of task “Charlie Chaplin”	44
4.11	Fact Sheet: Fire Alarm Story	45
4.12	Corpus uses of task “Fire Alarm Story”	45
4.13	Fact Sheet: Snowman Story	46
4.14	Corpus uses of task “Snowman Story”	46
4.15	Fact Sheet: Mr. Bean	46
4.16	Corpus uses of task “Mr. Bean”	47
4.17	Fact Sheet: Retelling of fables	47
4.18	Corpus uses of task “Retelling of fables”	47

4.19	Fact Sheet: Present yourself	48
4.20	Corpus uses of task “Present yourself”.	49
4.21	Fact Sheet: Sign Name	50
4.22	Corpus uses of task “Sign Name”.	50
4.23	Fact Sheet: Jokes	51
4.24	Corpus uses of task “Jokes”.	51
4.25	Fact Sheet: Free conversation	51
4.26	Corpus uses of task “Free conversation”.	52
4.27	Fact Sheet: Deaf life experiences	53
4.28	Corpus uses of task “Deaf life experiences”.	54
4.29	Fact Sheet: Your region	54
4.30	Corpus uses of task “Your region”.	55
4.31	Fact Sheet: Subject areas	55
4.32	Corpus uses of task “Subject areas”.	56
4.33	Fact Sheet: What did you do when it happened	58
4.34	Corpus uses of task “What did you do when it happened”.	58
4.35	Fact Sheet: Role play	59
4.36	Corpus uses of task “Role play”.	59
4.37	Fact Sheet: Volterra picture task	60
4.38	Corpus uses of task “Volterra picture task”.	60
4.39	Fact Sheet: Route description	60
4.40	Corpus uses of task “Route description”.	61
4.41	Fact Sheet: Calendar	62
4.42	Corpus uses of task “Calendar”.	62
4.43	Fact Sheet: Warning and prohibition signs	63
4.44	Corpus uses of task “Warning and prohibition signs”.	63
4.45	Fact Sheet: Describe process	64
4.46	Corpus uses of task “Describe process”.	64
4.47	Fact Sheet: Lexical elicitation	65
4.48	Corpus uses of task “Lexical elicitation”.	65
4.49	Fact Sheet: Diachronic changes	66
4.50	Corpus uses of task “Diachronic changes”.	66
4.51	Fact Sheet: Language awareness	67
4.52	Corpus uses of task “Language awareness”.	67
5.1	Fact Sheet: ASL Signbank	70
5.2	Fact Sheet: ASL-LEX	70
5.3	Fact Sheet: BSL SignBank	71
5.4	Fact Sheet: CDPSL	72
5.5	Fact Sheet: CZJ domain specific Lexicon	72
5.6	Fact Sheet: Danish Sign Language Dictionary	73
5.7	Fact Sheet: DGS Corpus types list	73
5.8	Fact Sheet: Dicta-Sign Lexicon	74
5.9	Fact Sheet: Dictio	75
5.10	Fact Sheet: DILSE	75
5.11	Fact Sheet: DIZLIS	76
5.12	Fact Sheet: DW-DGS	76
5.13	Fact Sheet: e-LIS	77
5.14	Fact Sheet: Elix	78
5.15	Fact Sheet: Finland-Swedish SignWiki	78

5.16 Fact Sheet: Finnish Signbank	79
5.17 Fact Sheet: Finnish SignWiki	79
5.18 Fact Sheet: GaLex	80
5.19 Fact Sheet: GLex	80
5.20 Fact Sheet: Global Signbank - LSFb	81
5.21 Fact Sheet: Global Signbank - NGT	82
5.22 Fact Sheet: Global Signbank - NTS	82
5.23 Fact Sheet: Global Signbank - VGT	83
5.24 Fact Sheet: HLex	83
5.25 Fact Sheet: LedaSila	84
5.26 Fact Sheet: Lex-LSFB	84
5.27 Fact Sheet: LSE-Sign	85
5.28 Fact Sheet: LSFb en ligne	85
5.29 Fact Sheet: NOEMA	86
5.30 Fact Sheet: Ocelles	86
5.31 Fact Sheet: PLex	87
5.32 Fact Sheet: POLYTROPON	87
5.33 Fact Sheet: sematos	88
5.34 Fact Sheet: SGB-FSS Lexicon	89
5.35 Fact Sheet: SignStudy	89
5.36 Fact Sheet: SLex	90
5.37 Fact Sheet: Slovar SZJ	90
5.38 Fact Sheet: SSLD	91
5.39 Fact Sheet: SUVI or Basic Dictionary of FinSL	91
5.40 Fact Sheet: Swiss German Sign Language Lexicon	92
5.41 Fact Sheet: TLex	92

ABBREVIATIONS

SL	sign language
ILI	interlingual lexical index
MT	machine translation
KWIC	keyword in context
HamNoSys	Hamburg Notation System for Sign Languages
LREC	International Conference on Language Resources and Evaluation
EMSL	European Meta-Sign Language
ASL	American Sign Language
BSL	British Sign Language
CZJ	Czech Sign Language
DGS	German Sign Language
DSGS	Swiss-German Sign Language
DTS	Danish Sign Language
FinSL	Finnish Sign Language
FinSSL	Finland-Swedish Sign Language
GSL	Greek Sign Language
HSL	Hungarian Sign Language
IS	International Sign
ISL	Irish Sign Language
LIS	Italian Sign Language
LSC	Catalan Sign Language
LSE	Spanish Sign Language
LSF	French Sign Language
LSFB	French Belgian Sign Language
NGT	Sign Language of the Netherlands
NTS	Norwegian Sign Language
ÖGS	Austrian Sign Language
PJM	Polish Sign Language
STS	Swedish Sign Language
SZJ	Slovene Sign Language
TİD	Turkish Sign Language
VGT	Flemish Sign Language



1 INTRODUCTION

EASIER mainly builds on two kinds of resources: linguistic corpora and broadcast data. While broadcast data are available in comparatively large quantity, linguistic corpora offer high quality data through rich transcription and linguistic annotation.

As technology to automatically tag or annotate sign language (SL) data in the quality required for linguistic annotation does not exist as of yet, corpus creators have met the challenge of having to manually annotate the data. Most of the SL corpus projects to date use ELAN (Crasborn and Sloetjes, 2008) or iLex (Hanke and Storz, 2008) for annotation. It is good practice in SL corpora to annotate SL data time-aligned to the source video, to base the annotations on a lexical database and to also offer time-aligned sentence-level translations for one or more spoken languages. Nonetheless different conventions for the notation of manual and non-manual activities are used throughout the different corpora. Especially the rich variety of non-lexical expressions and non-manual prosody (facial and torso signals) poses challenges to creating technologies for a large number of languages. Within EASIER, standardised notation conventions will be created and existing published corpora will be converted semi-automatically to conform to these standards.

This document lists the scope of data harmonisation candidates together with the most important features of these corpora from the perspective of using them as high-quality training data for the EASIER machine translation pipeline (see [Chapter 3](#)). Beyond the project, the document is most useful for researchers looking for existing SL corpora.

One of the goals of EASIER is to support researchers of under-resourced languages to develop SL technologies by sharing knowledge in form of best practices and guidelines harmonised for generation and annotation of new SL resources. To offer these guidelines, existing annotation conventions are compared. Patterns becoming apparent in the collection of linguistic corpora presented here can be used for the follow-up task of creating standardised notation conventions. Additionally, researchers from under-resourced languages are provided an overview of existing linguistic resources which they can browse through to get familiar with the existing approaches.

As multilingual signed corpora are so scarce, special attention is paid to elicitation tasks shared between corpora as this is what comes closest to multilingual corpora. In [Chapter 4](#) we identify elicitation tasks used in more than one project in identical format or being very close, and list the data based on these tasks.

Many SLs across Europe share parts of their grammar (including phonology) and some of their lexicon. The shared phonetic components make it possible to use one annotation system for many SLs without modification. The *Hamburg Notation System for Sign Languages* (*HamNoSys*) (Hanke, 2004) was developed at the Universität Hamburg to facilitate linguistic annotation. Many resources use HamNoSys for annotating the phonetic component.

One of the major challenges that EASIER faces is that not enough data exists for any individual SL to reach acceptable performance in machine translation (MT) and SL-recognition. To address this issue of data sparsity, the EASIER framework will combine data from multi-

ple European SLs to form a European Meta-Sign Language (EMSL) set of training data. This dataset will be used in intermediate steps of the EASIER pipeline to leverage the aforementioned commonalities between the SLs of Europe. The final steps of the pipeline will then convert the information from EMSL to the actual target SL. It must be stressed that EMSL is not a language in its own right and is in no way aimed at reducing the linguistic diversity across European SLs, but purely a technology-driven approach inspired by state-of-the-art MT solutions for less resourced spoken languages. In addition, EASIER aims to leverage existing linguistic annotation resources to provide a transferable phonetic representation and grammar that can be employed for less resourced and understudied European SLs in the future.

Furthermore EASIER will also link various European SLs on a concept level with an interlingual lexical index (ILI). The ILI will be used project-internally to drive lexical instantiation when going from EMSL to individual target SLs and contributes more generally to SL resources research.

This approach requires lexical resources for the target SLs. [Chapter 5](#) lists available lexical resources along with information on their linguistic annotation (e.g. HamNoSys annotation) which can be used for the EMSL and the ILI. Coverage will be increased over the course of the project.

The report is structured as follows: [Chapter 2](#) presents the methodology used to create the overview of datasets. [Chapters 3 to 5](#) each present different kinds of resources, presenting their specifics and a quick overview in the form of a factsheet. [Chapter 3](#) lists 26 corpora, [Chapter 4](#) presents 26 data collection tasks and [Chapter 5](#) contains 40 lexical resources. [Chapter 6](#) concludes the report.

2 METHODOLOGY

2.1 COLLECTION OF METADATA

To collect the metadata presented in the *Overview of datasets for the SLs of Europe* (henceforth: *Overview*) three main sources of information have been used: (1) the sign-lang@LREC anthology, (2) webpages assigned to the datasets or the projects which created the datasets, and (3) personal communication with data authors.

The *sign-lang@LREC anthology*¹ collects all papers published at the *Workshop Series on the Representation and Processing of Sign Languages*, which is held every two years in conjunction with the *International Conference on Language Resources and Evaluation (LREC)* as well as select papers from the LREC main conference that address sign languages. As such it collects a significant portion of the existing papers on linguistic resources for sign languages.

The examination of scientific papers was complemented with research done on the internet. Missing information was collected from webpages assigned to the datasets or their associated projects as well as from data repositories storing the datasets or metadata on these datasets.

Finally, to supplement the publicly available information through personal communication, a project-internal workshop was held, in which data authors within the EASIER consortium were informed about the method of metadata collection, the criteria applied and the encountered challenges. A first tabular version of the Overview was made available online via the EASIER shared drive for the project partners to review. Furthermore, data authors outside the EASIER consortium were contacted. They received our profile of their corpus and were invited to add missing information or even whole datasets as well as to correct outdated information.

In total we analysed 361 papers from 10 different LREC events identifying 137 datasets on 62 languages (signed and spoken). A few more data records were added through online research. To identify the candidates for data harmonisation as well as sources for the multilingual lexical resource to be built in EASIER and collection tasks used in several corpora as indicators on possible parallel data, threshold criteria as outlined in [Section 2.2](#) were defined. Applying these criteria, 66 datasets (26 corpora, 40 lexical resources) and 26 data elicitation tasks were included in the Overview and described in more detail. Some other datasets could not be included because they are not available to the public. We hope that cross-corpus research will become easier in the future when more projects offer their data with open access or at least make their metadata publicly available.

2.2 SELECTION CRITERIA

As described above, we set up criteria to collect the best candidates for different pipelines within EASIER. The criteria are described separately for corpora, data collection tasks and lexical resources.

¹<https://www.sign-lang.uni-hamburg.de/lrec/>

2.2.1 Corpora

We list linguistic corpora on SLs of Europe that contain spontaneous to semi-spontaneous signing (e. g. free signing and elicitation tasks that result in longer stretches or signing, but no simple repetition tasks). In general, the minimum size threshold applied was ten hours of signing, but multilingual resources and resources connected to a wordnet below this threshold were not filtered out. Following the idea that sign language corpora should be representative (see [Johnston, 2008](#)), we excluded corpora consisting of sign supported-only data or L2-beginners-only data.

2.2.2 Data Collection Tasks

We list data collection tasks if they are used for data elicitation in more than one of the corpora listed in [Chapter 3](#).

2.2.3 Lexical Resources

We list lexical resources in digital format if they cover SLs of Europe and appear to be useful within the EASIER framework, e. g. we excluded inaccessible resources and data on historical SLs. Resources connected to a wordnet did not have to meet these conditions.

2.2.4 Additional Information

By multilingual corpus we mean a corpus that contains more than one SL. Additional written languages used only for annotation and transcription do not count in this respect.

We would like to emphasise that [Chapter 5](#) in particular does not represent a complete collection of all possible lexical resources. As we learn of other useful resources, they will be added to the list.

3 RESOURCES: CORPORA

Sign language corpora consist of video recordings of sign language use, commonly enhanced by textual annotations. Unlike for many spoken languages, no computational tools exist yet to perform annotations (semi-)automatically. This makes the annotation of such corpora a labour-intensive manual task and as a result no larger sign language corpus is completely annotated. In addition, it is worth mentioning that there are no commonly agreed on annotation standards. A collection of linguistic corpora, as presented here, offers the possibility to compare annotation schemes to detect common patterns which could be used to create standardised notation conventions. With the Auslan Corpus Project², Johnston did pioneering work for the field of sign language corpus linguistics (Johnston, 2010). Its annotation guidelines are used (sometimes in a modified version) by various datasets. Depending on resources and research interests, the way and depth of annotation is adapted in each project.

The Overview contains the most important features of the corpora from the perspective of using them as high-quality training data for the EASIER machine translation pipeline. These include general information on the dataset, size of the dataset, composition of participants, data collection material used, kind and size of annotation, availability and access to the data and where to find it.

3.1 BRITISH SIGN LANGUAGE CORPUS

The British Sign Language Corpus is a collection of British Sign Language (BSL) video clips of 249 deaf signers from the UK. The BSL Corpus project is based at the Deafness Cognition and Language Research Centre, University College London, lasted from 2008–2011 and was led by Adam Schembri. A related dataset is the BSL Signbank (Section 5.3).

Metadata on the participants was collected via 39 questions on personal and language background following the standards for meta data collection by Crasborn and Hanke (2003) and using the IMDI format.

The recordings were made in a studio, using three cameras in three different angles (one on each signer and one on the pair). The participants were recorded in pairs, sitting next to each other in front of a blue background. They were asked in advance to wear plain coloured clothing. The tasks were moderated by a deaf researcher, except for the unobserved conversation mentioned above for which they would leave the room.

Table 3.1: *Fact Sheet: British Sign Language Corpus*

Name	British Sign Language Corpus
Language	BSL
Size	125 hours recorded, 70,000 tokens annotated
Participants	249 participants
	Deaf

²<https://www.auslan.org.au/about/corpus/>

Table 3.1: *Fact Sheet: British Sign Language Corpus (cont.)*

Name	British Sign Language Corpus
	4 age groups: 18–35, 35–50, 51–70 and 71 years and older
	From 8 cities: London, Bristol, Birmingham, Manchester, Newcastle, Glasgow, Cardiff, Belfast
	Balanced for gender, ethnicity, social class and language background
Metadata Format	IMDI
Translation	English, size unknown
Annotation	Annotation based on research projects (different bundles of annotation), basically following Johnston (2010)
Data Formats	ELAN
Licence	CC BY-SA 4.0 (for narrative and lexical elicitation data) User license (for conversation and interview data)
Access	Public access via browsable homepage Open access to narrative and lexical elicitation data Restricted access to conversation and interview data requires confirmed registration
Webpage	Project page: https://bslcorpusproject.org/ Dataset: https://bslcorpusproject.org/cava/ Public access: https://bslcorpusproject.org/data/region/
Institution	University College London
Publications	https://bslcorpusproject.org/publications-and-presentations/sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/bslcorpus.html

3.2 CORLSE

The Corpus de la Lengua de Signos Española (CORLSE) is a collection of Spanish Sign Language (LSE) video clips of 70 signers from Spain. CORLSE is based at the Centro de Normalización Lingüística de la Lengua de Signos Española (CNLSE). The creation of the corpus started 2014 and is ongoing.

The recordings were made in the premises of the associations of the deaf. The participants were recorded in pairs, sitting opposite each other in front of a blue background. A deaf interviewer was leading through the tasks.

Table 3.2: *Fact Sheet: CORLSE*

Name	CORLSE
Language	LSE
Size	<i>not available</i>

Table 3.2: Fact Sheet: CORLSE (cont.)

Name	CORLSE
Participants	70 participants
	4 age groups: 18–30, 31–50, 51–65 and 65 years and older
	From 4 regions: Madrid, Valencia, Granada, Valladolid
Metadata Format	<i>not available</i>
Translation	Spanish, size unknown
Annotation	Following the system of the majority of European corpora, together with the Australian (Johnston, 2010) and Catalan Sign Language corpus
Data Formats	ELAN
Licence	<i>not available</i>
Access	Public access via browsable homepage
	Restricted access for researchers, teachers and experts to annotations and sociolinguistic data requires confirmed registration
Webpage	Project page: https://corpuslse.es/
	Dataset: https://corpuslse.es/mapa.php
Institution	Centro de Normalización Lingüística de la Lengua de Signos Española (CNLSE)
Publications	<i>not available</i>

3.3 CORPUS LSFB

The LSFB Corpus is a collection of French Belgian Sign Language (LSFB) data from 100 signers from the Walloon Region of Belgium and Brussels. The LSFB Corpus project ran from 2012–2015, was based at the French Belgian Sign Language Laboratory (LSFB-Lab) at the University of Namur and led by Laurence Meurant.

A diverse set of data collection tasks was used to collect different discourse genres. Some of the tasks were adopted from the DGS Corpus (Section 3.9), NGT Corpus (Section 3.4), VGT Corpus (Section 3.6) and Auslan Corpus (Johnston and Schembri, 2006). A deaf moderator lead through the tasks

The signers came into a studio in pairs and were filmed with three JVC Pro HD 3 CCD cameras from two different angles: upper body and wide shot of both. The moderator was filmed with a Sony DV Handycam. The videos are in HD resolution with 50 frames per second.

Table 3.3: Fact Sheet: Corpus LSFB

Name	Corpus LSFB
Language	LSFB
Size	150 hours recorded, 104,000 tokens annotated

Table 3.3: *Fact Sheet: Corpus LSFB (cont.)*

Name	Corpus LSFB
Participants	100 participants
	30 native signers, 26 near-native signers, 44 late signers
	4 age groups: 18–25, 26–45, 46–65, 66–95 years old
	57 female, 43 male
Metadata Format	<i>not available</i>
Translation	French, 2.5 hours translated (1.6%)
Annotation	Following Johnston (2010) 12 hours annotated (8%)
Data Formats	ELAN, EDIUS, Lex-LSFB
Licence	CC BY-NC-SA 4.0
Access	Public access requires registration
	Restricted access for researchers, teachers, students and interpreters requires confirmed registration
Webpage	Project page including data: https://www.corpus-lsfb.be
Institution	University of Namur
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/corpuslsfb.html

3.4 CORPUS NGT

The Corpus NGT is an open access online corpus of dialogues between 100 native users of Sign Language of the Netherlands (NGT). Corpus NGT is based at Radboud University Nijmegen and was created by Onno Crasborn, Inge Zwitserlood and Johan Ros in a two-year project from 2006–2008. Different follow-up projects worked with the corpus and extended the amount of annotated data available online. The Corpus NGT team created their own annotation conventions (see Crasborn et al., 2020). A special feature is the voice-over translation for parts of the data, instead of the common translations into written Dutch, which also exists for other parts of the corpus.

Recordings are made with two HDV cameras and two digital MiniDV cameras. The participants were recorded in pairs, sitting opposite each other in front of a dark background. The signing is captured in a front view and a top view (from above). Recordings took place at the Radboud University and the Max Planck Institute for Psycholinguistics as well as at Deaf schools, Deaf clubs and other familiar places to the Deaf participants. A Deaf signer led the participants through the recordings.

Table 3.4: *Fact Sheet: Corpus NGT*

Name	Corpus NGT
Language	NGT

Table 3.4: Fact Sheet: Corpus NGT (cont.)

Name	Corpus NGT
Size	72 hours recorded, 150,000 tokens and 3,300 types annotated
Participants	100 participants
	Deaf, native signers
	8 age groups: 11–19, 20–29, 30–39, 40–49, 50–59, 60–69, 70–79, 80–89 years
	From 5 regions: Amsterdam, Groningen, Rotterdam, Gestel, Voorburg
Metadata Format	IMDI
Translation	Dutch, 15 hours translated, 15,000 sentences (21%)
Annotation	See Crasborn et al. (2020)
Data Formats	ELAN
Licence	Openly accessible videos under CC BY-NC-SA 3.0 NL
	Annotations under BY-NC-SA 4.0
Access	Public access via browsable homepage
	Open access to some video and annotation material via The Language Archive
	Restricted access to some video and metadata requires individual license agreement
Webpage	Project page: https://www.ru.nl/cls/our-research/research-groups/sign-language-linguistics/completed-projects/completed-projects/corpus-ngt/
	Dataset: https://hdl.handle.net/1839/00-0000-0000-0004-DF8E-6
	Public access: https://www.corpusngt.nl/
Institution	Radboud University Nijmegen
Publications	https://www.ru.nl/cls/our-research/research-groups/sign-language-linguistics/publications/ sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/corpusngt.html

3.5 CORPUS OF FINNISH SIGN LANGUAGE

The Corpus of Finnish Sign Language (CFinSL) is a collection of Finnish Sign Language (FinSL) and Finland-Swedish Sign Language (FinSSL) from 104 signers of Finland. The Corpus of FinSL is based at the University of Jyväskylä and was constructed under the lead of Tommi Jantunen in a five-year project from 2013–2018.

The recordings took place in a studio of the Audio-visual Research Centre filming two signers sitting opposite each other in front of a blue background. The recordings were made with seven HD cameras placed in five different angles. One camera filming the interviewer, one the total of the two signers, one the bird-view, one per signer filming in a 45-degree angle and one per

signer for a close-up of the upper body. A native signer is leading through the tasks.

Table 3.5: Fact Sheet: Corpus of Finnish Sign Language

Name	Corpus of Finnish Sign Language
Language	FinSL, FinSSL
Size	91 hours recorded (estimate), 107,000 tokens annotated
Participants	104 participants
	92 participants for FinSL
	12 participants for FinSSL
Metadata Format	IMDI
Translation	Finnish, size unknown
Annotation	ID-glosses, codes for lexicalised, depicting, gestural, numeral signs and fingerspelling, see Salonen et al. (2019) 15.25 hours (17%)
Data Formats	ELAN, SLMotion
Licence	CC BY-NC-SA 4.0
Access	Open access to 15.25 hours of video with annotation and translation
Webpage	Project page: https://www.jyu.fi/hytk/fi/laitokset/kivi/opiskelu/tutkinto-ohjelmat-ja-oppiaineet/viittomakieli/tutkimus-2/suomen-viittomakielten-korpusprojekti
	Dataset: https://korp.csc.fi/download/cfinsl/elicit/
	Meta share entry: http://urn.fi/urn:nbn:fi:lb-2019012321
Institution	University of Jyväskylä
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/cfinsl.html

3.6 CORPUS VLAAMSE GEBARENTAAL

The Corpus Vlaamse Gebarentaal is a collection of Flemish Sign Language (VGT) video material from 120 signers from the Flemish Region of Belgium. The project Corpus VGT is located at the University of Gent and ran from July 2012 to November 2015. The project was lead by Mieke Van Herreweghe and Myriam Vermeerbergen.

For the recordings signers came into a studio in pairs and were place sitting in front of a blue background. In total three cameras were used, one per signer and one for a total view. A moderator was leading through the tasks.

Table 3.6: Fact Sheet: Corpus Vlaamse Gebarentaal

Name	Corpus Vlaamse Gebarentaal
Language	VGT
Size	140 hours recorded, 5TB of data
Participants	120 participants
	6 age groups: 12–18, 19–25, 26–35, 36–50, 51–70, 71–99 years
	60 female, 60 male
	From 5 regions: Antwerpen, Limburg, Oost-Vlaanderen, Vlaams-Brabant, West-Vlaanderen
Metadata Format	<i>not available</i>
Translation	Dutch, size unknown
Annotation	<i>not available</i>
	<i>not available</i>
Data Formats	ELAN
Licence	CC BY-NC-SA 3.0
Access	Public access via browsable homepage
	Restricted access for teachers to more videos and annotation requires confirmed registration
	Restricted access for students to more videos requires confirmed registration
	Restricted access for researchers to all videos, annotation and metadata requires confirmed registration
Webpage	Project page: https://www.corpusvgt.be/
Institution	Universiteit Gent
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/corpusvgt.html

3.7 CREAGEST

The CREAGEST corpus is a corpus of adult and child French Sign Language (LSF) and of natural gestures. It consists of three sub corpora: a child acquisition dataset, a dataset of dialogues between deaf adults and a dataset of natural gestures. For the acquisition data 65 deaf children and 17 deaf adults were recorded by four deaf investigators. For the dialogue dataset 51 interviews were conducted by four deaf investigators. For the gestural dataset pairs of five hearing-hearing, five Deaf-Deaf and Deaf-hearing individuals were recorded. In total more than 500 hours of over 250 signers have been recorded. The Creagest project was based at the Centre national de la recherche scientifique (CNRS) at the Université Paris 8, ran from 2007–2012 and was led by Christian Cuxac.

To collect LSF production from the children four tasks – free conversations as well as controlled elicitation – were used. For the dialogues between deaf adults semi-directive interviews were conducted, followed by a metalinguistic discussion on the lexical units collected. For the dataset

of natural gestures the different pairs were presented to two explanation tasks.

Children were recorded with two cameras, adult interviews with three cameras. No information was found on the recording conditions of the gestural dataset.

Table 3.7: *Fact Sheet: CREAGEST*

Name	CREAGEST
Language	LSF
Size	500 hours recorded, 300 hours digitized
Participants	More than 250 participants
	Deaf and hearing
	Adults: 18–60 years old
	Children: 3–15 years old
	From 4 regions
Metadata Format	OLAC and IMDI
Translation	<i>not available</i>
Annotation	<i>not available</i> ~1 hour annotated
Data Formats	ELAN
Licence	CC BY-NC-ND 3.0
Access	Access to subset of videos via Ortolang requires registration
Webpage	Dataset dialogue: https://www.ortolang.fr/market/corpora/ortolang-000926
	Dataset acquisition: https://www.ortolang.fr/market/corpora/ortolang-000916
Institution	Centre national de la recherche scientifique (CNRS)
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/creagegstacquisition.html , https://www.sign-lang.uni-hamburg.de/lrec/data/creagegstdialogueentreadultessourds.html

3.8 DANISH SIGN LANGUAGE CORPUS

The Danish Sign Language Corpus is a collection of video material from 31 signers of DTS from Denmark. The Corpus is used to build a DTS-Danish Dictionary [Section 5.6](#). The Danish Sign Language Dictionary project building the corpus is based at the Bachelor's Degree Programme in Danish Sign Language and Speech-to-text Interpreter at the University College Copenhagen and led by Mads Jonathan Pedersen and Thomas Troelsgård. The project started 2014 and is still ongoing.

For the lexicographic work the Danish wordnet *DanNet*³ (Pedersen et al., 2009) was imple-

³<https://cst.ku.dk/english/projects/dannet/>

mented into the corpus.

The recordings took place in a classroom, in a rather informal setting. One camera was used filming the participants from a front view. The signers were recorded alone, having a deaf staff member as receiver, who also gave instructions on the tasks.

Table 3.8: *Fact Sheet: Danish Sign Language Corpus*

Name	Danish Sign Language Corpus
Language	Danish Sign Language (DTS)
Size	14.5 hours recorded, 20,000 tokens and 1,750 types annotated
Participants	31 participants
Metadata Format	Custom format
Translation	<i>not available</i>
Annotation	ID-glosses, mouthing, meaning in context annotated based on Cormier et al. (2017) and Crasborn et al. (2020) 2.5 hours (17%)
Data Formats	iLex
Licence	Individual license agreement for researchers
Access	No public access
Webpage	<i>not available</i>
Institution	University College Copenhagen
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/dtscorpus.html

3.9 DGS CORPUS

The DGS Corpus is a collection of German Sign Language (DGS) data from 330 signers from Germany. The 15-year long-term project is based at the Institute of German Sign Language and Communication of the Deaf at the Universität Hamburg and started in 2009. It is led by Thomas Hanke and Annika Herrmann. The DGS Corpus is used to build the DGS-German dictionary *DW-DGS* ([Section 5.12](#)).

The signers were recorded in pairs in a mobile studio travelling to 13 spots in Germany. The signers were sitting opposite each other in front of a blue background. In total seven cameras were used for the recordings, five HD cameras and two Bumblebees. The Bumblebees were later replaced by three HD stereo cameras. The cameras were set up in three different angles: one recording a total view including the moderator, one filming the signers from the front and one from above. The original resolution is 1080i50 for the videos of 2010 and 720p50 for the videos from 2011 onwards. Public data is provided in 360p50. A Deaf moderator was leading through the tasks.

Table 3.9: Fact Sheet: DGS Corpus

Name	DGS Corpus
Language	DGS
Size	560 hours recorded, 657,000 tokens annotated
Participants	330 participants
	4 age groups: 18–30, 31–45, 46–60, 60 years and older
	165 female, 165 male
	From all over Germany
Metadata Format	CMDI
Translation	German and English, 375.8 hours (German), 113 hours (English)
Annotation	See Konrad et al. (2020) 90.9 hours annotated
Data Formats	iLex
Licence	DGS Corpus License http://meine-dgs.de/ling/license_en.html
Access	Public access via browsable homepage
	Open access to 50 hours of video, annotation and translation in iLex, ELAN and SRT format
	Restricted access for researchers to further data requires individual license agreement
Webpage	Project page: http://dgs-korpus.de
	Dataset: http://ling.meine-dgs.de
	Public access: http://meine-dgs.de
Institution	Universität Hamburg
Publications	http://dgs-korpus.de/publications sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/dgscorpus.html

3.10 DICTA-SIGN CORPUS

The Dicta-Sign Corpus is a multilingual corpus of the four SLs BSL, DGS, Greek Sign Language (GSL) and LSF covering the topic of European travel. The corpus collects at least 14 informants per language. The recording sessions took approximately two hours and the same elicitation materials were used across languages. Dicta-Sign was a three-years project from the European's seventh framework programme. The consortium conducting the project consisted of eight partners: Institute for Language and Speech Processing, Universität Hamburg, University of East Anglia, University of Surrey, Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur (Limsi), Université Paul Sabatier, National Technical University of Athens, WebSourd.

In total 1,000 concepts with a SL equivalent in each language were collected (see [Section 5.8](#)) as well as training data for isolated signs and elicitation material for corpus collections.

Recordings were made with seven cameras recording the signers from different perspectives: front, side and top view as well as additional stereo cameras.

Table 3.10: Fact Sheet: Dicta-Sign Corpus

Name	Dicta-Sign Corpus
Language	BSL, DGS, GSL, LSF
Size	25 hours recorded, 1,000 types per language elicited
Participants	~60 participants
	14–16 participants per language
Metadata Format	IMDI
Translation	English, 0.6–5.5 hours translated per language
Annotation	Different bundles of annotation for different data sets containing segmentation, glosses, gaze, clause boundaries, HamNoSys <i>not available</i>
Data Formats	iLex
Licence	Individual license agreement for researchers
Access	Open access to subset of videos and elicitation material
Webpage	Project page: https://www.sign-lang.uni-hamburg.de/dicta-sign/portal/
Institution	Institute for Language and Speech Processing, Universität Hamburg, University of East Anglia, University of Surrey, Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur (LIMSI), Université Paul Sabatier, National Technical University of Athens, WebSourd
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/dictasigncorpus.html

3.11 DICTA-SIGN-GSL-V2

The Dicta-Sign-GSL-v2 is a remake of the GSL sub-corpus of the Dicta-Sign corpus (Section 3.10). It contains 5.5 hours of video material from 16 signers covering the topic of travel. The data was annotated in more detail, e.g. grammatical category on sentence and clause level. The Dicta-Sign-GSL-v2 was constructed at the Athena Research Center at the Institute for Language and Speech Processing (ILSP) from 2014–2016 and led by Eleni Efthimiou.

Recordings took place in a studio with two signers in each session sitting in front of a uni-coloured background. Six cameras were used for filming, two of them HD cameras and additional Kinect depth capturing cameras. The signers were filmed from three different angles: front, side and top view.

Table 3.11: Fact Sheet: Dicta-Sign-GSL-v2

Name	Dicta-Sign-GSL-v2
Language	GSL
Size	5.5 hours recorded
Participants	16 participants
Metadata Format	<i>not available</i>
Translation	English, 5.5 hours (100%) translated
Annotation	Lemmas, clause, boundaries, HamNoSys, classifiers, sentence type, clause type
	Fully annotated (100%)
Data Formats	iLex
Licence	<i>not available</i>
Access	<i>not available</i>
Webpage	<i>not available</i>
Institution	Institute for Language and Speech Processing, Athena Research Center
Publications	<i>not available</i>

3.12 DICTA-SIGN-LSF-V2

The Dicta-Sign-LSF-v2 is an extended version of the LSF sub-corpus of the Dicta-Sign corpus [Section 3.10](#) providing primary data (videos), elicitation data, annotation data and a related annotation guide, as well as preprocessed signer data including facial pose, upper body pose and hand shape estimates. It contains nine dialogue sessions with 18 signers of LSF covering the topic of travel. The data was annotated in more detail and a convolutional-recurrent learning network was trained on the data, drawing on a compact and generalisable modeling of the signers to provide a baseline for the recognition of lexical signs and non-lexical structures. Dicta-Sign-LSF-v2 was built at the Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur (LIMSI) in 2020.

Table 3.12: Fact Sheet: Dicta-Sign-LSF-v2

Name	Dicta-Sign-LSF-v2
Language	LSF
Size	11 hours recorded, 35,000 tokens annotated
Participants	16 participants
Metadata Format	<i>not available</i>
Translation	French, 11 hours (100%) translated
Annotation	Lexical sign, illustrative sign, pointing, tag, number, typing and gesture annotated
	Fully annotated (100%)

Table 3.12: Fact Sheet: *Dicta-Sign-LSF-v2* (cont.)

Name	Dicta-Sign-LSF-v2
Data Formats	Open Pose, CNN models
Licence	CC BY-NC-SA 4.0
Access	Access to videos, elicitation material and partial annotations via Ortolang requires registration
Webpage	Dataset: https://hdl.handle.net/11403/dicta-sign-lsf-v2/v1
Institution	Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur (LIMSI)
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/dictasignlsfv2.html

3.13 ECHO CORPUS

The European Cultural Heritage Online (ECHO) corpus is a multilingual corpus containing video material from three SLs: NGT, BSL and Swedish Sign Language (STS). Eight signers were recorded for 1.5 hours following the same tasks in each language. For NGT and BSL sign language poetry was added to the corpus. Additionally annotated segments of the *Gehörlos So!* corpus of DGS (Heßmann, 2001) were added to the corpus. The Echo project was a 18-month EU funded project dedicated to bring Essential Cultural Heritage online. The ECHO corpus was built from 2003–2004 by the Max Planck Institute for Psycholinguistics, Radboud University and University of Lund.

Filming took place in a studio with one or two signers at the same time. The signers were sitting or standing and depending on the task, recorded separately or closely next to each other. A single-coloured background was used.

Table 3.13: Fact Sheet: *ECHO Corpus*

Name	ECHO Corpus
Language	BSL, NGT, STS, DGS
Size	1.5 hours recorded
Participants	8 participants
	Native signers
	20–40 years old
Metadata Format	IMDI, OLAC
Translation	Dutch, English and Swedish, size unknown
Annotation	See Nonhebel et al. (2004)
Data Formats	ELAN
Licence	CC BY-NC-ND 3.0
Access	Open access to videos and transcripts via Language Archive
Webpage	Project page: http://sign-lang.ruhosting.nl/echo/

Table 3.13: *Fact Sheet: ECHO Corpus (cont.)*

Name	ECHO Corpus
	Dataset: https://hdl.handle.net/1839/00-0000-0000-0001-4892-C
Institution	Max Planck Institute for Psycholinguistics, Radboud University Nijmegen, University of Lund
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/echocorpus.html

3.14 GIVING COGNITION A HAND CORPUS

The Giving Cognition a Hand corpus is a multilingual corpus of Turkish Sign Language (TİD) and NGT as well as Turkish and Dutch data. It contains 84 video files of signers and speakers from Istanbul and Nijmegen. The project was based at the Max Planck Institute for Psycholinguistics, Centre for Language Studies.

Table 3.14: *Fact Sheet: Giving Cognition a Hand Corpus*

Name	Giving Cognition a Hand Corpus
Language	NGT, TİD
Size	84 video files
Participants	Number unknown
	Signers and speakers
	Adults and children
Metadata Format	CMDI
Translation	<i>not available</i>
Annotation	<i>not available</i>
Data Formats	<i>not available</i>
Licence	<i>not available</i>
Access	No public access
Webpage	Project page: https://www.nwo.nl/en/projects/277-70-013
	Dataset: https://hdl.handle.net/1839/bd27da7f-32e0-45bf-af16-c24a49fbf4d8
Institution	Max Planck Institute for Psycholinguistics, Centre for Language Studies
Publications	https://www.nwo.nl/en/projects/277-70-013

3.15 HUNGARIAN SIGN LANGUAGE CORPUS

The Hungarian Sign Language Corpus is a collection of Hungarian Sign Language (HSL) video data of 147 signers from Hungarian. All together 1,750 hours were recorded. The HSL corpus project ran from 2016–2017, was based at the Research Institute for Linguistics at the

Hungarian Academy of Sciences and led by Csilla Bartha.

Interviews were recorded with three cameras, while grammatical tests were filmed with five cameras.

Table 3.15: Fact Sheet: Hungarian Sign Language Corpus

Name	Hungarian Sign Language Corpus
Language	HSL
Size	1,750 hours recorded, 209 types annotated
Participants	147 participants
	Deaf
	21–82 years old
	75 female, 67 male
	From 9 regions: Budapest, Szeged, Hódmezővásárhely, Békéscsaba, Debrecen, Kaposvár, Sopron, Győr, Vác
Metadata Format	<i>not available</i>
Translation	Hungarian, 41 of 147 sociolinguistic interviews translated (estimate: 140 hours, 8%)
Annotation	140 tiers for annotation on all linguistic levels
Data Formats	ELAN
Licence	<i>not available</i>
Access	No public access
Webpage	<i>not available</i>
Institution	Hungarian Academy of Sciences
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/hslcorpus.html

3.16 IPROSLA CORPUS

The IPROSLA corpus is an archive bringing together two existing datasets of NGT on SL acquisition with the goal of providing documentation, metadata and long-term storage. One dataset consists of longitudinal data of deaf children from deaf and hearing parents collected at the University of Amsterdam over 20 years. The other dataset contains newly collected longitudinal data from hearing and deaf children of deaf parents collected by Radboud University. In total 16 children have been recorded for approximately 225 hours.

The recordings were done in informal settings like the home of the children. The children were interacting freely and playing with their parents and/or a deaf research assistant visiting the families.

For the recordings one to two cameras were used, originally with PAL mini-DV tapes or PAL VHS cassettes which were digitised later on.

Table 3.16: Fact Sheet: IPROSLA Corpus

Name	IPROSLA Corpus
Language	NGT
Size	225 hours recorded
Participants	16 child participants and their parents
	4 Deaf, 1 hard of hearing and 11 hearing
	6 months and older
Metadata Format	CMDI
Translation	<i>not available</i>
Annotation	<i>not available</i>
Data Formats	<i>not available</i>
Licence	Individual license agreement available
Access	Restricted access for researchers requires individual license agreement
Webpage	<p>Project page: https://www.ru.nl/cls/our-research/research-groups/sign-language-linguistics/completed-projects/completed-projects/iprosla-gebarengroei/</p> <p>Dataset: https://hdl.handle.net/1839/00-CE31E27F-8853-4A18-80E8-AECAFAD012C0</p> <p>Entry at Virtual Language Observatory: https://vlo.clarin.eu/record/https_58_47_47_hdl.handle.net_47_1839_47_00-CE31E27F-8853-4A18-80E8-AECAFAD012C0?1&count=980&index=0&q=iprosla</p>
Institution	University of Amsterdam, Radboud University Nijmegen
Publications	Crasborn (2010)

3.17 ITALIAN SIGN LANGUAGE CORPUS

The Italian Sign Language Corpus is a collection of Italian Sign Language (LIS) data from 180 signers of Italy. The core part of the project involved three universities: University of Milan-Bicocca, University Ca'Foscari and Sapienza University.

The data collection followed the main lines of the tasks used in the American Sign Language Corpus ([Lucas et al., 2002](#)) and the Auslan Corpus ([Johnston and Schembri, 2006](#)).

Signers were recorded in pairs or groups of three, sitting opposite each other with one camera filming each signer.

Table 3.17: Fact Sheet: Italian Sign Language Corpus

Name	Italian Sign Language Corpus
Language	LIS

Table 3.17: Fact Sheet: Italian Sign Language Corpus (cont.)

Name	Italian Sign Language Corpus
Size	100 hours recorded (estimate)
Participants	180 participants
	Native and later-exposed signers
	3 age groups: 18–30, 31–54, 55 years and older
	90 female, 90 male
Metadata Format	CMD
Translation	<i>not available</i>
Annotation	Annotation based on research projects (different bundles of annotation)
Data Formats	ELAN
Licence	CC BY-NC-SA
Access	Restricted access for researchers requires individual license agreement
Webpage	Dataset: https://hdl.handle.net/1839/00-57EA1164-AC96-4541-8B1D-252673D6152A
Institution	University of Milan-Bicocca, University Ca'Foscari, Sapienza University
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/lisocopus.html

3.18 MEDI-API-SKEL

MEDI-API-SKEL is a 2D-skeleton video corpus of LSF with French subtitles. The corpus consists of 368 subtitled videos produced by Média'Pi⁴, a media company producing bilingual content with LSF and written French. The corpus was produced at the Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur (LIMSI).

From the original 368 videos 135 body keypoints were extracted from every signer in every frame using OpenPose. The subtitles are time aligned to the 2D-skeleton video content. Skeleton videos in Json format and subtitles in WebVTT format are openly available, for the original videos an agreement with Média'Pi is needed.

Table 3.18: Fact Sheet: MEDI-API-SKEL

Name	MEDI-API-SKEL
Language	LSF
Size	27 hours recorded, 17,000 tokens from subtitles
Participants	more than 100 signers
Metadata Format	<i>not available</i>

⁴<https://media-pi.fr/>

Table 3.18: *Fact Sheet: MEDI-API-SKEL (cont.)*

Name	MEDI-API-SKEL
Translation	Fully translated (subtitles) (100%)
Annotation	Time aligned subtitles
Data Formats	Open Pose
Licence	CC BY-NC-SA 4,0 (partly) Open access to skeleton videos and subtitles
Access	Restricted access to video recordings for consultation only via Ortolang requires registration
	Further access requires individual license agreement
Webpage	Dataset: https://hdl.handle.net/11403/mediapi-skel/v1
Institution	Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur (LIMSI)
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/mediapiskel.html

3.19 PJM CORPUS

The Corpus of Polish Sign Language is a collection of video data from 150 Deaf native signers of Polish Sign Language (PJM). The PJM Corpus is based at the Laboratory of Sign Linguistics at the University of Warsaw and led by Paweł Rutkowski. The PJM Corpus project started in 2009 and is still ongoing. The project was able to win Trevor Johnston, creator of the Auslan Corpus, as an external consultant and advisor.

The participants are recorded in pairs in a recording studio. A deaf moderator is leading through the sessions. Tasks for collecting data were borrowed from other SL corpora, for example the DGS Korpus (Section 3.9). For recording HD cameras are used.

On basis of the Corpus of PJM the Corpus Dictionary of Polish Sign Language (Section 5.4), an online available for free dictionary for PJM including example sentences form the PJM Corpus, is created.

Table 3.19: *Fact Sheet: PJM Corpus*

Name	PJM Corpus
Language	PJM
Size	400 hours recorded, 505,000 tokens and 150,000 types annotated
Participants	150 participants
	Deaf
	Controlled for age, gender, region, age of acquisition, social background, education
Metadata Format	<i>not available</i>

Table 3.19: *Fact Sheet: PJM Corpus (cont.)*

Name	PJM Corpus
Translation	Polish, 10,000 clauses translated
Annotation	Segmentation, lemmatisation, tagging and articulation transcription in HamNoSys annotated 100,000 tokens already tagged for grammatical features
Data Formats	iLex, YouTrack
Licence	<i>not available</i>
Access	Not yet publicly available
Webpage	Project page: https://www.plm.uw.edu.pl/projekty/korpus-pjm/
Institution	University of Warsaw
Publications	https://www.plm.uw.edu.pl/publikacje/ sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/korpuspjm.html

3.20 POLYTROPON

The POLYTROPON parallel corpus is a corpus of GSL and Greek. The corpus consists of 3,600 sentences performed by a single signer in three repetitions each. The POLYTROPON corpus was constructed at the Athena Research Center at the Institute for Language and Speech Processing (ILSP) under the lead of Eleni Efthimiou.

Basis for the POLYTROPON parallel corpus is the POLYTROPON lexicon resource ([Section 5.32](#)). For each sign entry in the lexicon a GSL example of use was recorded and translated to Modern Greek. The annotation provides information for the grammar levels of lexicon, morphology, syntax and semantics.

Recordings were made with one HD and one kinect camera capturing the front view of the signer. The recording took place in a studio with uni-coloured background.

Table 3.20: *Fact Sheet: POLYTROPON*

Name	POLYTROPON
Language	GSL
Size	3,600 utterances with 3 repetitions each, 10,000 lexicon entries, 1,600 lemmas annotated
Participants	1 participant
Metadata Format	<i>not available</i>
Translation	Modern Greek, 3,500 sentences translated
Annotation	GR glosses, clause boundaries, HamNoSys, SiS-Builder non-manuals annotation tool, classifiers, sentence type and clause type fully annotated (100%)

Table 3.20: Fact Sheet: POLYTROPON (cont.)

Name	POLYTROPON
Data Formats	iLex, ELAN
Licence	CC BY-NC-SA 4.0
Access	Access to video and ELAN files requires registration
Webpage	Dataset: http://sign.ilsp.gr/signilsp-site/index.php/en/ppc/ Entry at clarin:el: http://hdl.handle.net/11500/ATHENA-0000-0000-4C77-6
Institution	Institute for Language and Speech Processing, Athena Research Center
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/polytropon.html

3.21 SIGNOR CORPUS

The SIGNOR Corpus of SZJ is a collection of Slovene Sign Language (SZJ) video data from 80 signers of Slovenia. The Corpus Signor project was based at the University of Ljubljana, ran from 2011–2014 and was led by Špela Vintar.

The annotation is based largely on the DGS Corpus Conventions (Konrad et al., 2020). Seven layers of annotation are provided: segmentation or tokenisation, glossing or lemmatisation, mouthing, HamNoSys transcription, Meaning, compositional meaning and segmentation into utterances. For the database of meanings the Slovene WordNet *SloWNet*⁵ (Fišer and Sagot, 2015) was adapted.

The recordings took place at the premises of Deaf clubs, partially at the informants homes and at the Deaf Institute Ljubljana. A moderator lead the participants through the tasks.

Table 3.21: Fact Sheet: SIGNOR Corpus

Name	SIGNOR Corpus
Language	SZJ
Size	40 hours recorded, 30,335 tokens and 1,976 types annotated
Participants	80 participants
Metadata Format	<i>not available</i>
Translation	<i>not available</i>
Annotation	Based on Konrad et al. (2020)
Data Formats	iLex
Licence	<i>not available</i>
Access	Public access via browsable homepage (temporarily unavailable at the time of writing)

⁵<http://lojze.lugos.si/darja/research/slownet/>

Table 3.21: *Fact Sheet: SIGNOR Corpus (cont.)*

Name	SIGNOR Corpus
Webpage	Project page: http://lojze.lugos.si/signor/en.html
Institution	University of Ljubljana
Publications	http://lojze.lugos.si/signor/en.html#objave sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/signorcorpus.html

3.22 SIGNS OF IRELAND

The Signs of Ireland Corpus is a collection of Irish Sign Language (ISL) video data from 40 signers of Ireland. The project was based at the Trinity College Dublin, took place in 2004 and was led by Lorraine Leesson.

Data collection was done by a deaf researcher, who is an established member of the Irish Deaf community. SL teachers were not recorded for the Corpus.

Table 3.22: *Fact Sheet: Signs of Ireland*

Name	Signs of Ireland
Language	ISL
Size	10 hours recorded (estimate)
Participants	40 participants
	Deaf, native and early signers
	4 age groups: 18–30, 30–45, 45–60, 65 years and older
	24 female, 16 male
	From 5 regions: Galway, Dublin, Wexford, Waterford, Cork
Metadata Format	<i>not available</i>
Translation	<i>not available</i>
Annotation	Following Nonhebel et al. (2004)
	all self selected stories and a quarter of the Frog stories
Data Formats	ELAN
Licence	<i>not available</i>
Access	No public access
Webpage	<i>not available</i>
Institution	Trinity College Dublin
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/signsofireland.html

3.23 SPANISH SIGN LANGUAGE CORPUS

The Spanish Sign Language Corpus is a collection of LSE video clips from 85 informants. The LSE corpus project is based at the University of Vigo.

Table 3.23: *Fact Sheet: Spanish Sign Language Corpus*

Name	Spanish Sign Language Corpus
Language	LSE
Size	<i>not available</i>
Participants	85 participants
Metadata Format	<i>not available</i>
Translation	Spanish, size unknown
Annotation	<i>not available</i>
Data Formats	ELAN
Licence	<i>not available</i>
Access	<i>not available</i>
Webpage	<i>not available</i>
Institution	Universidade de Vigo
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/lsecorpus.html

3.24 SWEDISH SIGN LANGUAGE CORPUS

The Swedish Sign Language Corpus is a collection of 42 signers from Sweden using STS. The STS Corpus project is based at Stockholm University, lasted from 2009–2011 and was led by Johanna Mesch.

The STS Corpus is one of a few corpora that provide part-of-speech tagging.

The signers were recorded in pairs, sitting opposite each other in a studio with a dark background. Five cameras were used for recording.

Table 3.24: *Fact Sheet: Swedish Sign Language corpus*

Name	Swedish Sign Language corpus
Language	STS
Size	24 hours recorded, 190,000 tokens and 18,800 types annotated
Participants	42 participants
	20–82 years old
	From three regions
Metadata Format	<i>not available</i>
Translation	Swedish, 14 hours translated (60%)

Table 3.24: Fact Sheet: Swedish Sign Language corpus (cont.)

Name	Swedish Sign Language corpus
Annotation	See Wallin and Mesch (2018), inspired by Nonhebel et al. (2004) and Johnston (2010) 24 hours transcribed (100%)
Data Formats	ELAN
Licence	CC BY-NC-SA
Access	Public access via browsable homepage Open access to video and parts of annotation data
Webpage	Project page: https://www.ling.su.se/teckenspraksresurser/teckensprakskorpusar/svensk-teckensprakskorpus Dataset: https://ling33.ling.su.se/sslc/video/ Public access: https://teckensprakskorpus.su.se
Institution	Stockholm University
Publications	https://www.ling.su.se/teckenspraksresurser/teckensprakskorpusar/svensk-teckensprakskorpus/publikationer sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/stskorpus.html

3.25 VID I SIGN SPACE CORPUS

The Vidi Sign Space Corpus is a corpus of DGS and TİD data collected by the Max Planck Institute for Psycholinguistics under the lead of Asli Özyürek from March 2007 to September 2012.

The signers were recorded in pairs in a studio sitting at a table. Three Sony DV cameras were used, one per signer and one for a top view.

Table 3.25: Fact Sheet: VID I Sign Space Corpus

Name	VID I Sign Space Corpus
Language	DGS, TİD
Size	135 hours of recording planned
Participants	<i>not available</i>
Metadata Format	IMDI, OLAC
Translation	<i>not available</i>
Annotation	Descriptive and analytic annotation
Data Formats	ELAN
Licence	<i>not available</i>
Access	Restricted access for researchers requires individual license agreement

Table 3.25: Fact Sheet: VIDI Sign Space Corpus (cont.)

Name	VIDI Sign Space Corpus
Webpage	Project page: https://www.nwo.nl/en/projects/276-70-009 Dataset: https://hdl.handle.net/1839/00-0000-0000-0008-68DB-A
Institution	Max Planck Institute for Psycholinguistics
Publications	https://www.nwo.nl/en/projects/276-70-009 sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/vidisignspacecorpus.html

3.26 VISIBASE CORPUS

The Visibase corpus is a collection of digitised and described NGT material that was present in the late 1990s at the sign language research groups at the University of Amsterdam and at Leiden University. The project lasted from 1996–2001 and was based at Radboud University, University of Amsterdam and Utrecht University.

Analogue video tapes were copied to professional digital video tapes (DVCAM). Metadata descriptions were created for all the data from Leiden and parts of the data from Amsterdam.

Table 3.26: Fact Sheet: Visibase Corpus

Name	Visibase Corpus
Language	NGT
Size	300 hours recorded
Participants	<i>not available</i>
Metadata Format	CMDI
Translation	<i>not available</i>
Annotation	<i>not available</i>
Data Formats	<i>not available</i>
Licence	<i>not available</i>
Access	Restricted access for researchers requires individual license agreement
Webpage	Dataset: https://hdl.handle.net/1839/00-0000-0000-0004-DF8F-4
Institution	Radboud University Nijmegen, University of Amsterdam, Utrecht University
Publications	<i>not available</i>

4 RESOURCES: DATA COLLECTION TASKS

To collect language data from signers and speakers, a number of different methods have been developed. Depending on the researchers' targeted discourse mode (e.g. narration, renarration, description, report, discussion, etc.) different material is used. For the collection of sign language data, visual stimuli are quite common as they minimise the influence of spoken languages (e.g. through written texts). Movies, pictures, photographs and other visual stimuli are used as well as signed questions and instructions.

To make the collected data more comparable we separately list common data collection tasks that have been used in the corpora presented in [Chapter 3](#). They can serve as a basis for cross-referencing resources in order to build near-parallel datasets. For each collection task the material and its possible usage is described. Next to a fact sheet describing the task itself we provide a listing of all resources using this task and where to find data online (if available).

Two remarks regarding the fact sheets: The *Degree of Interaction* is an estimate of ourselves regarding the interactivity of the task. *Duration* provides, where possible, an average duration of the participants' signing as observed in available data, otherwise an estimate is given.

4.1 SILVESTER AND TWEETY

“Canary Row” (Freleng, 1950) is a cartoon by Warner Bros. studios featuring Tweety the bird and Sylvester the cat. The cartoon is used widely by sign language researchers to elicit classifier constructions. The cartoon is shown to one of the participants, who then should describe the story to their dialogue partner. As this task is used within a lot of corpora the data can be used for cross-linguistic research.

Table 4.1: *Fact Sheet: Silvester and Tweety*

Name	Silvester and Tweety
Stimulus	Looney Tunes – Canary Row
Target	Data for cross-linguistic research
Degree of Interaction	Low (monologue)
Duration	10–15 min
Source	Freleng (1950), available at https://vimeo.com/317665278

Table 4.2: *Corpus uses of task “Silvester and Tweety”.*

Task	Silvester and Tweety
Resource (Language)	DGS Corpus (DGS)
# recordings – open access	3 (each in 7 parts)
# recordings – restricted access	81

Table 4.2: *Corpus uses of task “Silvester and Tweety”. (cont.)*

Task	Silvester and Tweety
Data available	http://meine-dgs.de/formats/format18_en.html
Resource (Language)	CORLSE (LSE)
# recordings – open access	15
# recordings – restricted access	<i>unknown</i>
Data available	https://corpuslse.es/busqueda_asociaciones_mapa.php?pagina=1&tarea=10 (Silvestre y Piolín)
Resource (Language)	PJM Corpus (PJM)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Resource (Language)	Corpus NGT (NGT)
# recordings – open access	46
# recordings – restricted access	7
Data available	https://hdl.handle.net/1839/00-0000-0000-0009-06F5-E
Resource (Language)	VIDI Sign Space Corpus (DGS)
# recordings – open access	0
# recordings – restricted access	112
Data available	https://hdl.handle.net/1839/00-0000-0000-0000-C71E-8
Resource (Language)	VIDI Sign Space Corpus (TID)
# recordings – open access	0
# recordings – restricted access	109
Data available	https://hdl.handle.net/1839/00-0000-0000-0008-430C-6

4.2 FROG STORY

“Frog, Where Are You?” (Mayer, 1969) is a picture book with no written language in it and is due to its parallel happening actions a popular task to elicit language data. Originally used in spoken language studies it was soon adopted by sign language researchers. The participants are given the book, or a digital version of the book is presented to them and they should retell the story to their dialogue partners. Due to its popularity this task collects data which can be used for cross-linguistic research.

Table 4.3: *Fact Sheet: Frog Story*

Name	Frog Story
Stimulus	Frog, where are you?

Table 4.3: Fact Sheet: Frog Story (cont.)

Name	Frog Story
Target	Data for cross-linguistic research
Degree of Interaction	Low (monologue)
Duration	5–10 min
Source	Mayer (1969), available at https://www.phil-fak.uni-duesseldorf.de/fileadmin/Redaktion/Institute/Allgemeine_Sprachwissenschaft/Frogstory-2_01.pdf

Table 4.4: Corpus uses of task “Frog Story”.

Task	Frog Story
Resource (Language)	CORLSE (LSE)
# recordings – open access	17
# recordings – restricted access	<i>unknown</i>
Data available	https://corpuslse.es/busqueda_asociaciones_mapa.php?pagina=1&tarea=5 (Rana, ¿dónde estás?)
Resource (Language)	Corpus NGT (NGT)
# recordings – open access	42
# recordings – restricted access	1
Data available	https://hdl.handle.net/1839/00-0000-0000-0009-06F9-3
Resource (Language)	SIGNOR Corpus (SZJ)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Resource (Language)	Corpus Vlaamse Gebarentaal (VGT)
# recordings – open access	37
# recordings – restricted access	<i>unknown</i>
Data available	https://www.corpusvgt.be/nl/corpussearch?field_geslacht_v1_value=All&field_geslacht_v1_value_1=All&field_leeftijd_v1_value=All&field_leeftijd_v1_value_1=All&field_regio_v1_value=All&field_regio_v1_value_1=All&field_thema_tid=10 (Kikker, waar ben je?)
Resource (Language)	DGS Corpus (DGS)
# recordings – open access	1 (in 6 parts)
# recordings – restricted access	81
Data available	http://meine-dgs.de/formats/format9_en.html
Resource (Language)	Corpus of Finnish Sign Language (FinSL, FinSSL)

Table 4.4: *Corpus uses of task “Frog Story”. (cont.)*

Task	Frog Story
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Resource (Language)	Corpus LSFB (LSFB)
# recordings – open access	0
# recordings – restricted access	22
Resource (Language)	PJM Corpus (PJM)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Resource (Language)	Signs of Ireland (ISL)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Resource (Language)	Swedish Sign Language corpus (STS)
# recordings – open access	23
# recordings – restricted access	<i>unknown</i>
Data available	https://ling33.ling.su.se/sslc/video/ (Var är du, grodan?)
Resource (Language)	VIDI Sign Space Corpus (TiD)
# recordings – open access	0
# recordings – restricted access	7
Data available	https://hdl.handle.net/1839/00-0000-0000-0008-42CA-0

4.3 PEAR STORY

The “Pear Story” (Chafe, 1980) is a widely used task to elicit language data. The Story was created for spoken languages first but soon adopted to elicit sign language data. It is a language-free six-minute film made at the University of California at Berkeley that is shown to participants, who should describe what happens in the movie. The collected data can be used for cross-linguistic research, as it is a common elicitation task within linguistics.

Table 4.5: *Fact Sheet: Pear Story*

Name	Pear Story
Stimulus	Pear story
Target	Data for cross-linguistic research
Degree of Interaction	Low (monologue)
Duration	5–10 min

Table 4.5: Fact Sheet: Pear Story (cont.)

Name	Pear Story
Source	Chafe (1980), available at http://chafe.faculty.linguistics.ucsb.edu/pearfilm.htm

Table 4.6: Corpus uses of task “Pear Story”.

Task	Pear Story
Resource (Language)	CORLSE (LSE)
# recordings – open access	2
# recordings – restricted access	<i>unknown</i>
Data available	https://corpuslse.es/busqueda_asociaciones_mapa.php?pagina=1&tarea=14 (Historia de la pera)
Resource (Language)	DGS Corpus (DGS)
# recordings – open access	1
# recordings – restricted access	82
Data available	http://meine-dgs.de/formats/format5_en.html
Resource (Language)	Signs of Ireland (ISL)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Resource (Language)	PJM Corpus (PJM)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>

4.4 SIGNS MOVIE

The movie “Signs” (Hughes, 2008) is a short movie without spoken language, the protagonists communicate by showing each other written English words on paper. The end of the movie is open about the hearing status of the female protagonist. The movie is approximately five minutes long. Both informants are asked to watch the movie and exchange on the content.

This task is expected to collect signs about feelings and love as well as assumptions. If needed subtitles for other languages were added.

Table 4.7: Fact Sheet: Signs Movie

Name	Signs Movie
Stimulus	Signs Movie
Target	Conversation and signs expressing love, feelings, assumptions
Degree of Interaction	High (exchange about topic)

Table 4.7: Fact Sheet: Signs Movie (cont.)

Name	Signs Movie
Duration	5–10 min
Source	Hughes (2008), available at https://www.youtube.com/watch?v=DzIOvv2ROE4 .

Table 4.8: Corpus uses of task “Signs Movie”.

Task	Signs Movie
Resource (Language)	DGS Corpus (DGS)
# recordings – open access	1
# recordings – restricted access	141
Data available	http://meine-dgs.de/formats/format20_en.html
Resource (Language)	PJM Corpus (PJM)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>

4.5 CHARLIE CHAPLIN

Charlie Chaplin is a well known actor in silent films. A movie – or scenes from his movies – is shown to the participants, who are asked to retell the story.

Table 4.9: Fact Sheet: Charlie Chaplin

Name	Charlie Chaplin
Stimulus	<i>not available</i>
Target	<i>not available</i>
Degree of Interaction	Low (monologue)
Duration	<i>not available</i>
Source	<i>not available</i>

Table 4.10: Corpus uses of task “Charlie Chaplin”.

Task	Charlie Chaplin
Resource (Language)	PJM Corpus (PJM)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Resource (Language)	VIDI Sign Space Corpus (DGS)
# recordings – open access	0

Table 4.10: *Corpus uses of task “Charlie Chaplin”. (cont.)*

Task	Charlie Chaplin
# recordings – restricted access	39
Data available	https://hdl.handle.net/1839/00-0000-0000-0009-2BB5-1
Resource (Language)	VIDI Sign Space Corpus (TiD)
# recordings – open access	0
# recordings – restricted access	42
Data available	https://hdl.handle.net/1839/00-0000-0000-0009-4513-5

4.6 FIRE ALARM STORY

The “Fire Alarm” (Matthes et al., 2010) is a short clip of a Deaf person telling about what happened during his last holiday trip to Italy. After the travel group arrived late at the hotel, he had a drink at the bar then went to bed and slept all night. The next morning the others tell him, that they tried to wake him, because the fire alarm went off during the night. The clip is shown to one of the informants, who then is asked to retell it to the other informant.

Aim of the task is to collect a high amount of sign language characteristic features, e. g. constructed action, nonmanuals, etc.

Table 4.11: *Fact Sheet: Fire Alarm Story*

Name	Fire Alarm Story
Stimulus	Signed Story
Target	Sign language characteristic features
Degree of Interaction	Low (monologue)
Duration	2–3 min
Source	See Matthes et al. (2010)

Table 4.12: *Corpus uses of task “Fire Alarm Story”.*

Task	Fire Alarm Story
Resource (Language)	DGS Corpus (DGS)
# recordings – open access	1
# recordings – restricted access	67
Data available	http://meine-dgs.de/formats/format7_en.html
Resource (Language)	PJM Corpus (PJM)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>

4.7 SNOWMAN STORY

“Snowman” (Briggs, 1978) is a picture book about a boy and a snowman. One of the informants is asked to look at the book and retell the story to the other informant.

Table 4.13: *Fact Sheet: Snowman Story*

Name	Snowman Story
Stimulus	Snowman Story
Target	<i>not available</i>
Degree of Interaction	Low (monologue)
Duration	<i>not available</i>
Source	Briggs (1978), available at https://www.arvindguptatoys.com/arvindgupta/snowman-eng.pdf

Table 4.14: *Corpus uses of task “Snowman Story”.*

Task	Snowman Story
Resource (Language)	Corpus of Finnish Sign Language (FinSL, FinSSL)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Resource (Language)	Swedish Sign Language corpus (STS)
# recordings – open access	21
# recordings – restricted access	<i>unknown</i>
Data available	https://ling33.ling.su.se/sslc/video/ (Snögubben)

4.8 MR. BEAN

Mr. Bean is a well known British sitcom with very little spoken language. Part of the series is shown to the participants, who are asked to retell the story.

Table 4.15: *Fact Sheet: Mr. Bean*

Name	Mr. Bean
Stimulus	Mr. Bean movie clip
Target	<i>not available</i>
Degree of Interaction	Low (monologue)
Duration	<i>not available</i>
Source	Davies et al. (1990–1995)

Table 4.16: *Corpus uses of task “Mr. Bean”.*

Task	Mr. Bean
Resource (Language)	Corpus of Finnish Sign Language (FinSL, FinSSL)
# recordings – open access	0
# recordings – restricted access	unknown
Resource (Language)	Swedish Sign Language corpus (STS)
# recordings – open access	21
# recordings – restricted access	unknown
Data available	https://ling33.ling.su.se/sslc/video/ (Film “Mr. Bean”)

4.9 RETELLING OF FABLES

One of the signers is presented to a fable and asked to retell it to the other signer. Commonly used fables are from the Grimm brother or Aesop, e. g. “*The tortoise and the hare*” (Perry 226), “*The shepherd’s boy and the wolf*”/“*The boy who cried wolf*” (Perry 210), “*The lion and the mouse*” (Perry 150), “*The two friends and the bear*” (Perry 65), “*The dog and his reflection*” (Perry 133).

Table 4.17: *Fact Sheet: Retelling of fables*

Name	Retelling of fables
Stimulus	Fables
Target	Data for cross-linguistic research
Degree of Interaction	Low (monologue)
Duration	1–2 min
Source	Available at https://aesopfables.com/

Table 4.18: *Corpus uses of task “Retelling of fables”.*

Task	Retelling of fables
Resource (Language)	Corpus NGT (NGT)
# recordings – open access	55
# recordings – restricted access	10
Data available	https://hdl.handle.net/1839/00-0000-0000-0009-06F7-8
Resource (Language)	Corpus Vlaamse Gebarentaal (VGT)
# recordings – open access	35 (tortoise and hare)
# recordings – restricted access	unknown

Table 4.18: *Corpus uses of task “Retelling of fables”. (cont.)*

Task	Retelling of fables
Data available	https://www.corpusvgt.be/nl/corpussearch?field_geslacht_v1_value=All&field_geslacht_v1_value_1=All&field_leeftijd_v1_value=All&field_leeftijd_v1_value_1=All&field_regio_v1_value=All&field_regio_v1_value_1=All&field_thema_tid=8 (Schildpad & haas)
Resource (Language)	ECHO Corpus (BSL)
# recordings – open access	10
# recordings – restricted access	0
Data available	https://hdl.handle.net/1839/00-0000-0000-0001-4950-1
Resource (Language)	ECHO Corpus (NGT)
# recordings – open access	20
# recordings – restricted access	0
Data available	https://hdl.handle.net/1839/00-0000-0000-0001-49C8-8
Resource (Language)	ECHO Corpus (STS)
# recordings – open access	10
# recordings – restricted access	0
Data available	https://hdl.handle.net/1839/00-0000-0000-0001-4AD9-1
Resource (Language)	Signs of Ireland (ISL)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>

4.10 PRESENT YOURSELF

Signers are asked to present themselves to each other. This task is good to let the signers get comfortable with the studio set-up.

Table 4.19: *Fact Sheet: Present yourself*

Name	Present yourself
Stimulus	Signed instruction
Target	Warum up
Degree of Interaction	Low (monologue)
Duration	0.5–2 min

Table 4.20: Corpus uses of task “Present yourself”.

Task	Present yourself
Resource (Language)	CORLSE (LSE)
# recordings – open access	17
# recordings – restricted access	<i>unknown</i>
Data available	https://corpuslse.es/busqueda_asociaciones_mapa.php?pagina=1&tarea=1 (Presentación)
Resource (Language)	Corpus NGT (NGT)
# recordings – open access	0
# recordings – restricted access	46
Data available	https://hdl.handle.net/1839/00-0000-0000-0009-06FA-5
Resource (Language)	SIGNOR Corpus (SZJ)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Resource (Language)	Swedish Sign Language corpus (STS)
# recordings – open access	22
# recordings – restricted access	<i>unknown</i>
Data available	https://ling33.ling.su.se/sslc/video/ (Presentation)
Resource (Language)	Visibase Corpus (NGT)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Resource (Language)	VIDI Sign Space Corpus (DGS)
# recordings – open access	0
# recordings – restricted access	20
Data available	https://hdl.handle.net/1839/00-0000-0000-0009-2B4E-5
Resource (Language)	VIDI Sign Space Corpus (TiD)
# recordings – open access	0
# recordings – restricted access	23
Data available	https://hdl.handle.net/1839/00-0000-0000-0009-6FC4-0

4.11 SIGN NAME

For this task the participants are asked to show their sign name and explain where it comes from. Sign names are an inherent part of Deaf culture and therefore interesting to collect. This task works as well as a warm up for the participants.

Table 4.21: Fact Sheet: Sign Name

Name	Sign Name
Stimulus	Signed instruction
Target	Documentation of Deaf culture, warm up
Degree of Interaction	Low (monologue)
Duration	2–3 min

Table 4.22: Corpus uses of task “Sign Name”.

Task	Sign Name
Resource (Language)	Corpus Vlaamse Gebarentaal (VGT)
# recordings – open access	36
# recordings – restricted access	<i>unknown</i>
Data available	https://www.corpusvgt.be/nl/corpussearch?field_geslacht_v1_value=All&field_geslacht_v1_value_1=All&field_leeftijd_v1_value=All&field_leeftijd_v1_value_1=All&field_regio_v1_value=All&field_regio_v1_value_1=All&field_thema_tid=12 (Naamgebaar)
Resource (Language)	DGS Corpus (DGS)
# recordings – open access	0
# recordings – restricted access	168
Resource (Language)	Corpus LSFB (LSFB)
# recordings – open access	0
# recordings – restricted access	50
Resource (Language)	PJM Corpus (PJM)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Resource (Language)	Swedish Sign Language corpus (STS)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>

4.12 JOKES

The informants are asked to tell each other a joke. Similar to the Sign name task (Section 4.12) this task captures a part of Deaf culture as Deaf jokes are a part of it and it is a good warm up. If this task is prepared by the participants at home the collected data is expected to be a prepared signing monologue.

Table 4.23: *Fact Sheet: Jokes*

Name	Jokes
Stimulus	Ask for preparation at home
Target	Documentation of Deaf culture, warm up
Degree of Interaction	Low (monologue)
Duration	2–7 min

Table 4.24: *Corpus uses of task “Jokes”.*

Task	Jokes
Resource (Language)	DGS Corpus (DGS)
# recordings – open access	88
# recordings – restricted access	49
Data available	http://meine-dgs.de/formats/format21_en.html
Resource (Language)	Corpus LSFB (LSFB)
# recordings – open access	0
# recordings – restricted access	16
Resource (Language)	PJM Corpus (PJM)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Resource (Language)	Visibase Corpus (NGT)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>

4.13 FREE CONVERSATION

A lot of corpora offer free conversation episodes. Some of these recordings arise spontaneous in-between tasks, others are more planned. For the latter participants are asked to converse freely without being moderated. In some projects attention was paid, that the moderator leaves the room for this task. The task is ideal to collect spontaneous signing.

Table 4.25: *Fact Sheet: Free conversation*

Name	Free conversation
Stimulus	No stimulus
Target	Free (unobserved) signing
Degree of Interaction	High (dialogue)
Duration	15–30 min

Table 4.26: Corpus uses of task “Free conversation”.

Task	Free conversation
Resource (Language)	British Sign Language Corpus (BSL)
# recordings – open access	452
# recordings – restricted access	452
Data available	https://digital-collections.ucl.ac.uk/R/HN97JGFGQ94BPR12C88YTSJ3G95CJ4VVC4SMY9AF28RV8BJ49-00692?func=collections-result&collection_id=2648
Resource (Language)	Corpus NGT (NGT)
# recordings – open access	60
# recordings – restricted access	0
Data available	https://hdl.handle.net/1839/00-0000-0000-0009-06F8-6
Resource (Language)	Corpus Vlaamse Gebarentaal (VGT)
# recordings – open access	2
# recordings – restricted access	<i>unknown</i>
Data available	https://www.corpusvgt.be/nl/corpussearch?field_geslacht_v1_value=All&field_geslacht_v1_value_1=All&field_leeftijd_v1_value=All&field_leeftijd_v1_value_1=All&field_regio_v1_value=All&field_regio_v1_value_1=All&field_thema_tid=18 (Vrije conversatie)
Resource (Language)	DGS Corpus (DGS)
# recordings – open access	34
# recordings – restricted access	131
Data available	http://meine-dgs.de/formats/format8_en.html
Resource (Language)	Italian Sign Language Corpus (LIS)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Resource (Language)	Corpus of Finnish Sign Language (FinSL, FinSSL)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Resource (Language)	IPROSLA Corpus (NGT)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Data available	https://hdl.handle.net/1839/00-CE31E27F-8853-4A18-80E8-AECAFAD012C0
Resource (Language)	Corpus LSFb (LSFB)

Table 4.26: *Corpus uses of task “Free conversation”. (cont.)*

Task	Free conversation
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Resource (Language)	PJM Corpus (PJM)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Resource (Language)	Swedish Sign Language corpus (STS)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Resource (Language)	Visibase Corpus (NGT)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Resource (Language)	VIDI Sign Space Corpus (DGS)
# recordings – open access	0
# recordings – restricted access	5
Data available	https://hdl.handle.net/1839/00-0000-0000-0009-38E3-3
Resource (Language)	VIDI Sign Space Corpus (TiD)
# recordings – open access	0
# recordings – restricted access	23
Data available	https://hdl.handle.net/1839/00-0000-0000-0009-6FB4-5 , https://hdl.handle.net/1839/00-0000-0000-0009-4B5B-9

4.14 DEAF LIFE EXPERIENCES

Many SL corpora contain a task to document typical experiences of Deaf people, such as Deaf schools and education, Deaf retirement homes, associations of the Deaf, Deaf sports clubs, and so on. Participants are asked to share their experiences from Deaf schools, residential schools, Deaf retirement homes, Deaf sports clubs, associations of the Deaf, etc. The collected data is expected to be spontaneous and lively.

Table 4.27: *Fact Sheet: Deaf life experiences*

Name	Deaf life experiences
Stimulus	Pictures, comic strips or no material
Target	Documentation of Deaf culture
Degree of Interaction	Low (monologue)
Duration	5–15 min

Table 4.27: *Fact Sheet: Deaf life experiences (cont.)*

Name	Deaf life experiences
Source	Possibly: That Deaf Guy, available at https://thatdeafguy.com/

Table 4.28: *Corpus uses of task “Deaf life experiences”.*

Task	Deaf life experiences
Resource (Language)	Corpus NGT (NGT)
# recordings – open access	60
# recordings – restricted access	3
Data available	https://hdl.handle.net/1839/00-0000-0000-0009-06F6-0
Resource (Language)	DGS Corpus (DGS)
# recordings – open access	63
# recordings – restricted access	259
Data available	http://meine-dgs.de/formats/format3_en.html
Resource (Language)	Corpus LSFB (LSFB)
# recordings – open access	0
# recordings – restricted access	76 (deafclub), 61 (childhood), 50 (hobbies)
Resource (Language)	Swedish Sign Language corpus (STS)
# recordings – open access	12
# recordings – restricted access	unknown
Data available	https://ling33.ling.su.se/sslc/video/ (Döv ...)
Resource (Language)	PJM Corpus (PJM)
# recordings – open access	0
# recordings – restricted access	unknown

4.15 YOUR REGION

The informants are asked to talk about the specialities of the region they live in. Examples named are typical culinary specialities, sites, landscapes, products, customs, etc. In the DGS Corpus (Section 3.9) signers were grouped by region to facilitate the exchange on this topic. The aim of the task is to collect signs for names of places, famous happening, etc. and a discourse type of text.

Table 4.29: *Fact Sheet: Your region*

Name	Your region
Stimulus	Signed instruction

Table 4.29: *Fact Sheet: Your region (cont.)*

Name	Your region
Target	Signs for names of places and famous happenings
Degree of Interaction	High (discourse)
Duration	20 min
Source	See Nishio et al. (2010)

Table 4.30: *Corpus uses of task “Your region”.*

Task	Your region
Resource (Language)	DGS Corpus (DGS)
# recordings – open access	13
# recordings – restricted access	67
Data available	http://meine-dgs.de/formats/format10_en.html
Resource (Language)	PJM Corpus (PJM)
# recordings – open access	0
# recordings – restricted access	unknown

4.16 SUBJECT AREAS

Different subject areas are presented to the signers with pictures, written subject titles or in signed instructions. Commonly used subject areas are family, home, work, hobbies, childhood, education and travel. The signers are asked to describe the topics or their personal family, work, home, etc. This task aims at collecting a solid basis of basic vocabulary additionally the Deaf culture is documented. Some projects collect discussions on this task afterwards and provide them bundled together across different tasks.

Table 4.31: *Fact Sheet: Subject areas*

Name	Subject areas
Stimulus	Written phrases, pictures and/or signed instruction
Target	Basic vocabulary, documentation of Deaf culture
Degree of Interaction	Low – middle
Duration	5–15 min
Source	See for example Nishio et al. (2010)

Table 4.32: *Corpus uses of task “Subject areas”.*

Task	Subject areas
Resource (Language)	CORLSE (LSE)
# recordings – open access	14 (childhood), 16 (education), 15 (work)
# recordings – restricted access	<i>unknown</i>
Data available	https://corpuslse.es/busqueda_asociaciones_mapa.php?pagina=1&tarea=3 (Childhood – Anécdotas de la infancia); https://corpuslse.es/busqueda_asociaciones_mapa.php?pagina=1&tarea=2 (Education – Experiencia educativa); https://corpuslse.es/busqueda_asociaciones_mapa.php?pagina=1&tarea=4 (Work – Trabajo y aficiones)
Resource (Language)	Corpus Vlaamse Gebarentaal (VGT)
# recordings – open access	3 (education)
# recordings – restricted access	<i>unknown</i> (home, family, hobby)
Data available	https://www.corpusvgt.be/nl/corpussearch?field_geslacht_v1_value=All&field_geslacht_v1_value_1=All&field_leeftijd_v1_value=All&field_leeftijd_v1_value_1=All&field_regio_v1_value=All&field_regio_v1_value_1=All&field_thema_tid=11 (Schooltijd)
Resource (Language)	DGS Corpus (DGS)
# recordings – open access	26 (25 subject areas)
# recordings – restricted access	349
Data available	http://meine-dgs.de/formats/format13_en.html
Resource (Language)	Dicta-Sign Corpus (BSL)
# recordings – open access	8
# recordings – restricted access	8
Data available	https://www.sign-lang.uni-hamburg.de/dicta-sign/portal/task.html
Resource (Language)	Dicta-Sign Corpus (DGS)
# recordings – open access	7
# recordings – restricted access	7
Data available	https://www.sign-lang.uni-hamburg.de/dicta-sign/portal/task.html
Resource (Language)	Dicta-Sign Corpus (GSL)
# recordings – open access	8
# recordings – restricted access	8

Table 4.32: Corpus uses of task “Subject areas”. (cont.)

Task	Subject areas
Data available	https://www.sign-lang.uni-hamburg.de/dicta-sign/portal/task.html
Resource (Language)	Dicta-Sign Corpus (LSF)
# recordings – open access	8
# recordings – restricted access	8
Data available	https://www.sign-lang.uni-hamburg.de/dicta-sign/portal/task.html
Resource (Language)	Dicta-Sign-GSL-v2 (GSL)
# recordings – open access	0
# recordings – restricted access	8
Resource (Language)	Dicta-Sign-LSF-v2 (LSF)
# recordings – open access	8
# recordings – restricted access	8
Data available	https://hdl.handle.net/11403/dicta-sign-lsf-v2/v1
Resource (Language)	Corpus of Finnish Sign Language (FinSL, FinSSL)
# recordings – open access	0 (work, hobby)
# recordings – restricted access	<i>unknown</i>
Resource (Language)	Spanish Sign Language Corpus (LSE)
# recordings – open access	0 (family)
# recordings – restricted access	<i>unknown</i>
Resource (Language)	Corpus LSFB (LSFB)
# recordings – open access	0
# recordings – restricted access	11 (family), 18 (education)
Resource (Language)	VIDI Sign Space Corpus (DGS)
# recordings – open access	0
# recordings – restricted access	17 (family, house)
Data available	https://hdl.handle.net/1839/00-0000-0000-0008-AC09-5 (House), https://hdl.handle.net/1839/00-0000-0000-0008-ACOB-9 (Family)
Resource (Language)	VIDI Sign Space Corpus (TiD)
# recordings – open access	0
# recordings – restricted access	32 (house), 3 (family), 1 (room)

Table 4.32: *Corpus uses of task “Subject areas”. (cont.)*

Task	Subject areas
Data available	https://hdl.handle.net/1839/00-0000-0000-0009-6FBC-6 (House — collection 1), https://hdl.handle.net/1839/00-0000-0000-0008-4439-6 (House — collection 2), https://hdl.handle.net/1839/00-0000-0000-0008-444B-5 (Room), https://hdl.handle.net/1839/00-0000-0000-0008-4432-9 (Family)
Resource (Language)	Visibase Corpus (NGT)
# recordings – open access	0
# recordings – restricted access	<i>unknown (work)</i>
Resource (Language)	PJM Corpus (PJM)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>

4.17 WHAT DID YOU DO WHEN IT HAPPENED

Informants are presented to shocking or moving events in the past, e.g. the moon landing, the nuclear accident in Chernobyl, 9/11 or the death of Princess Diana and asked to report how they felt and what they did when they heard about the event. The description is signed and accompanied with pictures to evoke memories. Aim of the task is a lively description in monologues or dialogues and the collection of information on how deaf people experienced the events with their limited access to information.

Table 4.33: *Fact Sheet: What did you do when it happened*

Name	What did you do when it happened
Stimulus	Signed instruction and pictures
Target	Lively signing, documentation of Deaf culture
Degree of Interaction	Middle (can be monologues and dialogues)
Duration	20 min
Source	See Nishio et al. (2010)

Table 4.34: *Corpus uses of task “What did you do when it happened”.*

Task	What did you do when it happened
Resource (Language)	DGS Corpus (DGS)
# recordings – open access	52
# recordings – restricted access	279
Data available	http://meine-dgs.de/formats/format1_en.html

Table 4.34: *Corpus uses of task “What did you do when it happened”. (cont.)*

Task	What did you do when it happened
Resource (Language)	PJM Corpus (PJM)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>

4.18 ROLE PLAY

The participants are asked to imagine they have the opportunity to meet the mayor or a minister. They should convince them about some actions related to the Deaf community that should or should not happen. A list of actions is presented, e. g. the closing of Deaf clubs or interpreting in public TV. The participants have to prepare their arguments and present them to each other. The other signer can give advice or their opinion in the end.

Table 4.35: *Fact Sheet: Role play*

Name	Role play
Stimulus	Signed instruction
Target	Debate, reasoning
Degree of Interaction	Middle (monologue with comments)
Duration	10–15 min

Table 4.36: *Corpus uses of task “Role play”.*

Task	Role play
Resource (Language)	Corpus LSFB (LSFB)
# recordings – open access	0
# recordings – restricted access	41
Resource (Language)	Visibase Corpus (NGT)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>

4.19 VOLTERRA PICTURE TASK

A signer describes a picture to the other signer, who then have to select the right picture having the choice between two different ones. The stimuli consist of a series of 18 sets of paired pictures showing a series of situations that aim to elicit transitive utterances.

Table 4.37: *Fact Sheet: Volterra picture task*

Name	Volterra picture task
Stimulus	18 sets of paired pictures
Target	Transitive utterances
Degree of Interaction	Middle (instructing each other)
Duration	<i>not available</i>
Source	See Volterra et al. (1984)

Table 4.38: *Corpus uses of task “Volterra picture task”.*

Task	Volterra picture task
Resource (Language)	Corpus Vlaamse Gebarentaal (VGT)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Resource (Language)	Signs of Ireland (ISL)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Resource (Language)	VIDI Sign Space Corpus (DGS)
# recordings – open access	0
# recordings – restricted access	14
Data available	https://hdl.handle.net/1839/00-0000-0000-0008-AC0C-C
Resource (Language)	VIDI Sign Space Corpus (TID)
# recordings – open access	0
# recordings – restricted access	14
Data available	https://hdl.handle.net/1839/00-0000-0000-0009-4526-0

4.20 ROUTE DESCRIPTION

For the Route description task one participant is asked to describe the correct route on a city map to the other participant. The stimulus consists of two maps of the same place with different landmarks. The task should elicit among other things data on locations and projections of 2D maps into the signing space.

Table 4.39: *Fact Sheet: Route description*

Name	Route description
Stimulus	Two maps of the same place with different landmarks

Table 4.39: Fact Sheet: Route description (cont.)

Name	Route description
Target	Locations
Degree of Interaction	Middle (monologues dialogue)
Duration	10 min
Source	See Matthes et al. (2010)

Table 4.40: Corpus uses of task "Route description".

Task	Route description
Resource (Language)	Corpus Vlaamse Gebarentaal (VGT)
# recordings – open access	5
# recordings – restricted access	unknown
Data available	https://www.corpusvgt.be/nl/corpussearch?field_geslacht_v1_value=All&field_geslacht_v1_value_1=All&field_leeftijd_v1_value=All&field_leeftijd_v1_value_1=All&field_regio_v1_value=All&field_regio_v1_value_1=All&field_thema_tid=20 (Wegbeschrijving)
Resource (Language)	DGS Corpus (DGS)
# recordings – open access	1
# recordings – restricted access	65
Data available	http://meine-dgs.de/formats/format15_en.html
Resource (Language)	Dicta-Sign Corpus (BSL)
# recordings – open access	0
# recordings – restricted access	8
Resource (Language)	Dicta-Sign Corpus (DGS)
# recordings – open access	0
# recordings – restricted access	7
Resource (Language)	Dicta-Sign Corpus (GSL)
# recordings – open access	0
# recordings – restricted access	8
Resource (Language)	Dicta-Sign Corpus (LSF)
# recordings – open access	0
# recordings – restricted access	8
Resource (Language)	Corpus LSFB (LSFB)
# recordings – open access	0
# recordings – restricted access	50

Table 4.40: *Corpus uses of task “Route description”. (cont.)*

Task	Route description
Resource (Language)	PJM Corpus (PJM)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Data available	https://hdl.handle.net/1839/00-0000-0000-0000-C7D4-6
Resource (Language)	VIDI Sign Space Corpus (DGS)
# recordings – open access	0
# recordings – restricted access	10
Resource (Language)	VIDI Sign Space Corpus (TiD)
# recordings – open access	0
# recordings – restricted access	1
Data available	https://hdl.handle.net/1839/00-0000-0000-0008-4448-0

4.21 CALENDAR

Each informant gets a one-week calendar with fictitious appointments and they are asked to arrange two meetings of two hours each to prepare a surprise for the wedding party of a mutual friend. They are also told explicitly to talk about the other activities they have planned in that week. The aim is to collect a dialogue of negotiation and signs for the days of the week, time terms and various common activities such as seeing the doctor, going on vacation, being at work, sports activities, having a plumber at home or going to the movies and the theatre.

Table 4.41: *Fact Sheet: Calendar*

Name	Calendar
Stimulus	One-week calendar with fictive appointments
Target	Dialogue (planning and negotiation) and days of the week, time terms, activities etc.
Degree of Interaction	High (negotiating)
Duration	10 min
Source	See Nishio et al. (2010)

Table 4.42: *Corpus uses of task “Calendar”.*

Task	Calendar
Resource (Language)	DGS Corpus (DGS)
# recordings – open access	1

Table 4.42: *Corpus uses of task “Calendar”. (cont.)*

Task	Calendar
# recordings – restricted access	167
Data available	http://meine-dgs.de/formats/format16_en.html
Resource (Language)	PJM Corpus (PJM)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>

4.22 WARNING AND PROHIBITION SIGNS

Unusual warning and prohibition signs from all around the world are shown to the participants. They are invited to discuss what the signs could possibly mean. This task is a good warm up for the participants. Target of this task is to elicit negated sentences in a coherent context. 16 warning and prohibition signs are used in the DGS Corpus (Section 3.9).

Table 4.43: *Fact Sheet: Warning and prohibition signs*

Name	Warning and prohibition signs
Stimulus	Signed instruction and pictures
Target	Warm up, negation in context
Degree of Interaction	High (discussion with disagreement)
Duration	15 min
Source	See Nishio et al. (2010)

Table 4.44: *Corpus uses of task “Warning and prohibition signs”.*

Task	Warning and prohibition signs
Resource (Language)	Corpus Vlaamse Gebarentaal (VGT)
# recordings – open access	36
# recordings – restricted access	<i>unknown</i>
Resource (Language)	DGS Corpus (DGS)
# recordings – open access	16
# recordings – restricted access	152
Data available	http://meine-dgs.de/formats/format14_en.html
Resource (Language)	PJM Corpus (PJM)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>

4.23 DESCRIBE PROCESS

Signers are presented with some kind of instruction, cooking recipes, DIY instruction, IKEA assembly or common knowledge tasks like changing a tire. The signers describe the instructions to each other step-by-step. Goal of this task is to collect detailed descriptions and explanations of a sequence of actions as well as phrases to structure a text.

Table 4.45: Fact Sheet: Describe process

Name	Describe process
Stimulus	Step-by-step instructions, e. g. IKEA assembly, DIY picture instruction
Target	Description and explanation of actions, text structure
Degree of Interaction	Low (monologue)
Duration	5 min

Table 4.46: Corpus uses of task “Describe process”.

Task	Describe process
Resource (Language)	Corpus Vlaamse Gebarentaal (VGT)
# recordings – open access	3
# recordings – restricted access	0
Data available	https://www.corpusvgt.be/nl/corpussearch?field_geslacht_v1_value=All&field_geslacht_v1_value_1=All&field_leeftijd_v1_value=All&field_leeftijd_v1_value_1=All&field_regio_v1_value=All&field_regio_v1_value_1=All&field_thema_tid=19 (Beschrijven van procedures)
Resource (Language)	DGS Corpus (DGS)
# recordings – open access	13
# recordings – restricted access	153
Data available	http://meine-dgs.de/formats/format4_en.html
Resource (Language)	Corpus LSFb (LSFB)
# recordings – open access	0
# recordings – restricted access	58
Resource (Language)	PJM Corpus (PJM)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>

4.24 LEXICAL ELICITATION

Signers are asked to sign isolated signs, sometimes with an additional short explanation sentence of the sign. The signs are presented in written form and/or with pictures. The signs cover areas as colour, weekdays, numerals, seasons, the human body and others. The DGS Corpus (Section 3.9) elicited 34 terms, the BSL Corpus (Section 3.1) 102 concepts. This task is mainly focused on the study of (regional) variation.

Table 4.47: *Fact Sheet: Lexical elicitation*

Name	Lexical elicitation
Stimulus	Written word and/or picture
Target	Lexical variation
Degree of Interaction	Low (monologue)
Duration	10–15 min

Table 4.48: *Corpus uses of task “Lexical elicitation”.*

Task	Lexical elicitation
Resource (Language)	British Sign Language Corpus (BSL)
# recordings – open access	386
# recordings – restricted access	386
Data available	https://digital-collections.ucl.ac.uk/R/HN97JGFGQ94BPR12C88YTSJ3G95CJ4VVC4SMY9AF28RV8BJ49-00698?func=collections-result&collection_id=2651
Resource (Language)	DGS Corpus (DGS)
# recordings – open access	0
# recordings – restricted access	168
Resource (Language)	CORLSE (LSE)
# recordings – open access	16
# recordings – restricted access	<i>unknown</i>
Data available	https://corpuslse.es/busqueda_asociaciones_mapa.php?pagina=1&tarea=8 (Cuerpo humano y colores)
Resource (Language)	PJM Corpus (PJM)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Resource (Language)	ECHO Corpus (BSL)
# recordings – open access	1
# recordings – restricted access	0

Table 4.48: *Corpus uses of task “Lexical elicitation”. (cont.)*

Task	Lexical elicitation
Data available	https://hdl.handle.net/1839/00-0000-0000-0001-49AF-B
Resource (Language)	ECHO Corpus (NGT)
# recordings – open access	4
# recordings – restricted access	0
Data available	https://hdl.handle.net/1839/00-0000-0000-0001-4A68-0
Resource (Language)	ECHO Corpus (STS)
# recordings – open access	1
# recordings – restricted access	0
Data available	https://hdl.handle.net/1839/00-0000-0000-0001-4AE2-C

4.25 DIACHRONIC CHANGES

Informants are asked to sign about signs which are different between young and old generations. This task elicits meta-linguistic discourse as well as sociolinguistic variation.

Table 4.49: *Fact Sheet: Diachronic changes*

Name	Diachronic changes
Stimulus	Signed instruction
Target	Sociolinguistic variation, meta-linguistic discourse
Degree of Interaction	High (discussion)
Duration	5–10 min

Table 4.50: *Corpus uses of task “Diachronic changes”.*

Task	Diachronic changes
Resource (Language)	CORLSE (LSE)
# recordings – open access	16
# recordings – restricted access	<i>unknown</i>
Data available	https://corpuslse.es/busqueda_asociaciones_mapa.php?pagina=1&tarea=17 (Debate diacronía en LSE)
Resource (Language)	Corpus Vlaamse Gebarentaal (VGT)
# recordings – open access	2
# recordings – restricted access	<i>unknown</i>

Table 4.50: *Corpus uses of task “Diachronic changes”.* (cont.)

Task	Diachronic changes
Data available	https://www.corpusvgt.be/nl/corpussearch?field_geslacht_v1_value=All&field_geslacht_v1_value_1=All&field_leeftijd_v1_value=All&field_leeftijd_v1_value_1=All&field_regio_v1_value=All&field_regio_v1_value_1=All&field_thema_tid=28 (Oude vs. nieuwe gebaren)
Resource (Language)	DGS Corpus (DGS)
# recordings – open access	1
# recordings – restricted access	157
Data available	http://meine-dgs.de/formats/format19_en.html
Resource (Language)	Italian Sign Language Corpus (LIS)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>
Resource (Language)	PJM Corpus (PJM)
# recordings – open access	0
# recordings – restricted access	<i>unknown</i>

4.26 LANGUAGE AWARENESS

This task can have different instructions. In the BSL corpus (Section 3.1) participants were asked for their definition of BSL and about their opinion on variation and change, BSL teaching and so on. In the LSFb corpus (Section 3.3) the moderator asked the participants what “good” or “bad” signing is. In another task the participants were asked to discuss variations in LSFb between younger and older signers, interpreters and different regions. In the Corpus NGT (Section 3.4) participants were asked about Sign Language Issues.

Table 4.51: *Fact Sheet: Language awareness*

Name	Language awareness
Stimulus	Interview questions
Target	Meta-linguistic discourse
Degree of Interaction	Middle (interview or discussion)
Duration	15 min

Table 4.52: *Corpus uses of task “Language awareness”.*

Task	Language awareness
Resource (Language)	British Sign Language Corpus (BSL)
# recordings – open access	0

Table 4.52: *Corpus uses of task “Language awareness”. (cont.)*

Task	Language awareness
# recordings – restricted access	367
Data available	https://digital-collections.ucl.ac.uk/R/1N1I7A9DP4V65Y2LEYRFGQ3PNHEGD8I2BYD113KI6IGX52B6FP-08527?func=collections-result&collection_id=2649
Resource (Language)	Corpus NGT (NGT)
# recordings – open access	43
# recordings – restricted access	2
Data available	https://hdl.handle.net/1839/00-0000-0000-0009-06FF-7
Resource (Language)	Corpus LSFB (LSFB)
# recordings – open access	0
# recordings – restricted access	54 (emotions), 44 (norms and signing)

5 RESOURCES: LEXICAL RESOURCES

Our collection of lexical resources includes lexical databases as well as electronic dictionaries. A lexical database is a language resource containing lexemes and additional information on these lexemes. Lexical databases are primarily built as a tool to support the consistent annotation of corpus data, but they are also essential in dictionary creation as dedicated dictionary writing software does not yet exist for sign language data. Glosses are stored within the lexical database, together with further information such as phonological and morphological descriptions, translation equivalents, semantic relations and more, and can be used for glossing the corpus data. While lexical databases are sometimes used as dictionaries, caution is advised, as sign glosses are only very approximate translations that (1) cover only one of possibly several senses of the sign, (2) cannot specify which senses of the spoken-language gloss word actually apply to the sign, and (3) lack most of the additional information that a dictionary could provide.

Commonly used software for the creation of SL lexical databases are iLex and Signbank. iLex is an annotation tool with integrated lexical database that offers lexicon building while transcribing corpus data (Hanke, 2002). iLex was created at the Universität Hamburg, based on longstanding work on empirically based lexicographic and corpus analysis. It has been used as the basis for several lexical resources. Signbank is a web based lexical database software for sign languages originating from the Auslan Corpus project (Johnston, 2008). It has since been used by several other projects which extended it to fit their needs, resulting in multiple independent implementations (so called forks) written in Python and JavaScript⁶.

Dictionaries share with lexical databases that they contain lexemes and further information, such as citation forms or translations. Unlike lexical databases, dictionaries are written for public users and not (only) for researchers and may contain more detailed information, e. g. on sign usage or sense disambiguation. Dictionaries nowadays are often website-based. Within the EASIER framework, the lexical resources will be used to build a multi-SL lexical resource with an interlingual lexical index (ILI) as the backbone for use in the multilingual automatic translation process. The ILI will semantically align European SLs on the lexical level via concepts. To build the ILI, openly available lexical resources for all project languages are needed. The ILI will support the lexical instantiation step in the machine translation pipeline. The Overview lists the most important features of the lexical resources that are relevant for the integration into a multilingual lexical resource. The listed features include sources, institutions, size and content of the resources. It should be noted, however, that sizes are not directly comparable between the lexical resources as varying degrees of aggregation are used.

5.1 ASL SIGNBANK

ASL Signbank is a collection of ID glosses used for annotation in the Sign Language Acquisition: Annotation, Archiving, and Sharing (SLAAAASH) project. It is a consistent and constantly upgraded resource, that serves for ongoing annotations. ASL Signbank is connected to ASL-LEX (Section 5.2), a lexical database that includes iconicity and subjective frequency judgements.

⁶Most Signbank forks can be found at <https://github.com/Signbank>

ASL Signbank – together with ASL-LEX (Section 5.2) and SignStudy (Section 5.35) – is used in the construction of *ASLNet* (C. P. Lualdi et al., 2021), a wordnet for American Sign Language (ASL).

Table 5.1: Fact Sheet: ASL Signbank

Name	ASL Signbank
Language	ASL, English
Size	3,702 signs
Linguistic information	ID-gloss, citation form, translations to English, handedness
Licence	CC BY-NC-SA 4.0
Access	Public access to 2,848 signs via browsable homepage, more detailed information and all signs need confirmed registration
Webpage	https://aslsignbank.haskins.yale.edu/
Institution	SLAAASh Project, Haskins Lab, Yale University
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/aslsignbank.html

5.2 ASL-LEX

ASL-LEX is a database of phonological and lexical properties of ASL signs. When first released it consisted of 1,000 signs, growing in size to 2,723 signs in the 2020 release. ASL-LEX is built in a collaboration of Laboratory for Language and Cognitive Neuroscience at San Diego State University, the Programs in Deaf Studies at Boston University and the Psycholinguistics and Linguistics Lab at Tufts University. The creators are Naomi Caselli, Karen Emmorey, Zed Sevcikova Sehyr, Ariel Cohen-Goldberg, Cindy O’Grady Farnady.

The resource is available as a searchable web interface and as a raw data in spreadsheet form. The project provides web visualisations, instructions of usage and a download option.

The frequency ratings provided are done by 25 deaf signers per sign.

ASL-LEX – together with ASL Signbank (Section 5.1) and SignStudy (Section 5.35) – is used in the construction of *ASLNet* (C. P. Lualdi et al., 2021), a wordnet for ASL.

Table 5.2: Fact Sheet: ASL-LEX

Name	ASL-LEX
Language	ASL, English
Size	2,723 signs
Linguistic information	Frequency, iconicity and transparency ratings, phonological coding, neighborhood density, phonotactic probability, cross-references, visualizations, English translations for 25%
Licence	CC BY-NC 4.0

Table 5.2: Fact Sheet: ASL-LEX (cont.)

Name	ASL-LEX
Access	Public access via browsable homepage, open access to download raw data
Webpage	https://asl-lex.org/index.html
Institution	San Diego State University, Boston University and Tufts University
Publications	https://asl-lex.org/publications.html

5.3 BSL SIGNBANK

The BSL Signbank is a dictionary built from two sources: the BSL Corpus (Section 3.1) and the British Sign Language/English Dictionary (Brien, 1992). The two resources overlap to a large extent, but not completely. The BSL Signbank was built at the Deafness, Cognition and Language Research Centre (DCAL) at the University College London and last updated mid-2014.

BSL Signbank includes the signs used to annotate conversation data from Bristol, Birmingham, London and Manchester and those of all signers from the task “What’s your sign” (see Section 4.24). Further signs from more regions and tasks are planned to be added. User of the dictionary can also report missing signs which they would like to be added.

Table 5.3: Fact Sheet: BSL SignBank

Name	BSL SignBank
Language	BSL, English
Size	2,500 signs
Linguistic information	Citation form of variants, english translations; after login: Sign Number, ID gloss, Annotation ID Gloss, Phonology, Definitions and Notes
Licence	<i>not available</i>
Access	Public access via browsable homepage, for more detailed information confirmed registration needed
Webpage	https://bslsignbank.ucl.ac.uk/
Institution	University College London
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/bslsignbank.html

5.4 CDPSL

The Corpus-based Dictionary of Polish Sign Language (CDPSL) is a dictionary of PJM/Polish based on the PJM Corpus (Section 3.19). The dictionary and corpus are built at the Pracownia Lingwistyki Migowej at the University of Warsaw under the lead of Paweł Rutkowski.

The dictionary includes all PJM signs that appeared in the PJM Corpus more than 4 times as well as additional signs to fill obvious gaps in that. Signs are searchable by their form and certain semantic properties.

Table 5.4: *Fact Sheet: CDPSL*

Name	CDPSL
Language	PJM, Polish
Size	<i>not available</i>
Linguistic information	Citation form, HamNoSys, definition, example sentence, linked signs, types of use (instead of POS), handshape, localization, number of hands
Licence	<i>not available</i>
Access	Public access via browsable homepage
Webpage	https://www.slownikpjm.uw.edu.pl/en
Institution	University of Warsaw
Publications	https://www.plm.uw.edu.pl/publikacje/

5.5 CZJ DOMAIN SPECIFIC LEXICON

The domain-specific lexicon for Czech Sign Language (CZJ) for the area of train travels was built on the basis of the Czech to Signed Czech (CSC) parallel corpus (Kanis et al., 2006) and translations of train announcements. The lexicon was built at the Department of Cybernetics, Faculty of Applied Sciences, University of West Bohemia under the lead of Jakub Kanis.

Table 5.5: *Fact Sheet: CZJ domain specific Lexicon*

Name	CZJ domain specific Lexicon
Language	CZJ
Size	330 signs
Linguistic information	Domain of train travel, HamNoSys transcripts, animations
Licence	<i>not available</i>
Access	Not online available at the time of writing
Webpage	<i>not available</i>
Institution	University of West Bohemia
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/czjdomainspecificlexicon.html

5.6 DANISH SIGN LANGUAGE DICTIONARY

The Danish Sign Language dictionary is a corpus-based dictionary for DTS and Danish, publicly available via a browsable homepage. The dictionary was developed at the Centre for Sign

Language, University College Copenhagen under the lead of Thomas Troelsgård and in close cooperation with the Danish Deaf Association (DDL). Signs can be searched by handshape, location, Danish keywords and topics.

Table 5.6: *Fact Sheet: Danish Sign Language Dictionary*

Name	Danish Sign Language Dictionary
Language	DTS, Danish
Size	2,000 signs
Linguistic information	Citation form, translation to Danish, collocations, meaning, definitions, homophony, phonological variation
Licence	<i>not available</i>
Access	Public access via browsable homepage
Webpage	http://www.tegnsprog.dk/
Institution	University College Copenhagen
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/dtsdictionary.html

5.7 DGS CORPUS TYPES LIST

The DGS Corpus types list is a list of types extracted from the lexical database within iLex used to annotate video recordings in the DGS Corpus (Section 3.9). The publicly available types list covers all types that occur in the public part of the DGS Corpus.

Types within the DGS Corpus project are organised hierarchical in a parent-child relationship: each parent type (or token) is specified by a citation form and every child type (subtype) stand for a conventionalised form-meaning relation. The subtypes inherit their citation form and iconic value from the parent type. The naming of type glosses gives a hint to the iconic value of the sign whereas subtype glosses, like keywords, express a core meaning aspect. The typelist is arranged on the basis of the types. The citation form of each type is provided as HamNoSys transcription and (where available) as a video recording, shown from up to four camera angles. These videos are taken from DW-DGS (Section 5.12) and the DGS specialist dictionaries (Sections 5.18, 5.19, 5.24, 5.31 and 5.41). Underneath this, all transcript occurrences of the type are shown in a keyword in context (KWIC) view, grouped by subtype.

Table 5.7: *Fact Sheet: DGS Corpus types list*

Name	DGS Corpus types list
Language	DGS, German, English
Size	14,064 types
Linguistic information	Citation (filmed from four angles), ID-gloss, HamNoSys transcriptions, context KWIC view from the DGS Corpus transcripts, DOI
Licence	DGS Corpus license http://meine-dgs.de/ling/license_en.html

Table 5.7: Fact Sheet: DGS Corpus types list (cont.)

Name	DGS Corpus types list
Access	Public access via browsable homepage
Webpage	http://meine-dgs.de/ling/types_en.html
Institution	Universität Hamburg
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/dgscorpustypeslist.html

5.8 DICTA-SIGN LEXICON

Dicta Sign Lexicon is a multilingual lexicon for BSL, GSL, DGS, LSF, English, Greek, German and French. Approximately 1,000 concepts are provided for each of the project SLs. The shared list of concepts chosen for the lexicon is of everyday use or specifically related to the field of European travel. The lexicon provides recordings of each sign for each language, annotation with gloss labels, form description (HamNoSys) and a rough meaning. Dicta-Sign was a three-years project from the European's seventh framework programme. The consortium conducting the project consisted of eight partners: Institute for Language and Speech Processing, Universität Hamburg, University of East Anglia, University of Surrey, Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur (Limsi), Université Paul Sabatier, National Technical University of Athens, WebSourd.

The project also created the Dicta-Sign Corpus (see [Section 3.10](#)), training data for isolated signs and elicitation material for corpus collections.

Table 5.8: Fact Sheet: Dicta-Sign Lexicon

Name	Dicta-Sign Lexicon
Language	BSL, GSL, DGS, LSF, English, Greek, German, French
Size	1,000 signs per language
Linguistic information	Citation form, HamNoSys, translation to English, Greek, German and French
Licence	Individual license agreement for researchers
Access	Public access via browsable homepage; some videos temporarily unavailable at the time of writing
Webpage	https://www.sign-lang.uni-hamburg.de/dicta-sign/portal/
Institution	nstitute for Language and Speech Processing, Universität Hamburg, University of East Anglia, University of Surrey, Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur (LIMS), Université Paul Sabatier, National Technical University of Athens, WebSourd
Publications	<i>not available</i>

5.9 DICTIO

Dictio is a multilingual dictionary for Czech Sign Language (CZJ), Slovene Sign Language (SZJ), Austrian Sign Language (ÖGS), American Sign Language (ASL), International Sign (IS), Czech, Slovenian, German and English. Dictio is based at the Masaryk University. Signs can be searched via their form and translation keywords. Dictio is gradually growing and some datasets and search functions were not yet available at the time of writing.

Table 5.9: Fact Sheet: Dictio

Name	Dictio
Language	CZJ, SZJ, ÖGS, ASL, IS, Czech, Slovak, German, English
Size	<i>not available</i>
Linguistic information	Citation form, grammar information, etymology, stylistic information, HamNoSys or SignWriting transcription, semantic field, meaning, example of use, semantically superordinate and subordinate, synonymic and antonymic expressions, phraseological units, translations
Licence	<i>not available</i>
Access	Public access via browsable homepage, more content requires confirmed registration; keyword search temporarily unavailable at the time of writing
Webpage	https://www.dictio.info/
Institution	Masaryk University
Publications	https://www.dictio.info/about

5.10 DILSE

Dilse is a LSE-Spanish Dictionary produced at the Fundación CNSE. The dictionary was created with deaf professionals and deaf specialists in LSE. Videos and pictures of the signs can be shared online and are downloadable.

Table 5.10: Fact Sheet: DILSE

Name	DILSE
Language	LSE, Spanish
Size	9,000 signs
Linguistic information	Translation to Spanish, definition, picture, citation form
Licence	CC BY-NC-SA 3.0
Access	Public access via browsable homepage, content can be shared and downloaded
Webpage	https://fundacioncnse-dilse.org/
Institution	Fundación CNSE

Table 5.10: Fact Sheet: DILSE (cont.)

Name	DILSE
Publications	<i>not available</i>

5.11 DIZLIS

Dizlis is a LIS-Italian dictionary created by the Cooperativa Alba. The videos of Dizlis are also used in the dictionary e-LIS. The webpage offers searching signs by parameters as well as Italian keywords.

Table 5.11: Fact Sheet: DIZLIS

Name	DIZLIS
Language	LIS, Italian
Size	1,000 signs
Linguistic information	Examples of usage, significant phrases, synonyms, variants, parameters, pictures, in-depth notes
Licence	Dizlis license http://www.dizlis.it/web/termini-e-condizioni.html
Access	Public access via browsable homepage
Webpage	http://www.dizlis.it/
Institution	Cooperativa Alba
Publications	<i>not available</i>

5.12 DW-DGS

The Digitales Wörterbuch DGS (DW-DGS) is a corpus-based dictionary of DGS/German. The dictionary is built by the DGS Corpus project at the Institute of German Sign Language (IDGS), Universität Hamburg. Its content is primarily based on the DGS Corpus (Section 3.9), but also uses information collected through the online survey DGS-Feedback (Wähl et al., 2018) and the elicitation tool SignHunter (Hanke et al., 2020).

At the time of writing creation of the dictionary was still ongoing, so only a selection of preliminary entries is shown and the layout is not yet finalised. The dictionary is regularly updated with more signs. Signs can be searched by a visual graph, IDs or animated images, German keywords or subject groups.

Table 5.12: Fact Sheet: DW-DGS

Name	DW-DGS
Language	DGS, German
Size	250 entries comprising of 1,348 signs

Table 5.12: Fact Sheet: DW-DGS (cont.)

Name	DW-DGS
Linguistic information	ID, citation form, senses, regional distribution, related signs, similar signs, homonyms, antonyms and authentic samples, linked to the KWIC view of transcripts in the corpus
Licence	DW-DGS License http://dw-dgs.de/overview/intro.html#lizenz
Access	Public access via browsable homepage
Webpage	http://dw-dgs.de
Institution	Universität Hamburg
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/dwdgs.html

5.13 E-LIS

The Electronic Bilingual Dictionary of Italian Sign Language and Italian, was created at the European Academy of Bolzano. Signs can be searched via parameters or Italian keywords. e-LIS draws on the video material already present in the Dizlis dictionary (Section 5.11).

Table 5.13: Fact Sheet: e-LIS

Name	e-LIS
Language	LIS, Italian
Size	<i>not available</i>
Linguistic information	Definition, example, usage information and variants
Licence	<i>not available</i>
Access	Public access via browsable homepage, videos temporarily not available at the time of writing
Webpage	http://elisdiz.eurac.edu/diz/
Institution	Eurac Research
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/elis.html

5.14 ELIX

Elix is a dictionary for LSF/French working like a search engine. French keywords can be searched, hits show associated signs and their definition in LSF. Elix is built by Signes de Sens. Elix can be used as an online web platform and as an application.

Table 5.14: Fact Sheet: Elix

Name	Elix
Language	LSF, French
Size	15,300 signs
Linguistic information	Citation form, definitions
Licence	<i>not available</i>
Access	Public access via browsable homepage
Webpage	https://dico.elix-lsf.fr/
Institution	Signe de sens
Publications	<i>not available</i>

5.15 FINLAND-SWEDISH SIGNWIKI

The Finland-Swedish SignWiki is a multifaceted open access dictionary of FinSSL. It used crowdsourcing to collect the displayed information. Signs and other information can be uploaded and discussed online. A related SignWiki for FinSL also exists (see [Section 5.17](#)). SignWiki was created in a cooperation between the Finnish Association of the Deaf, the Humanities University of Applied Sciences (Humak) and the Icelandic Communication Center for the Deaf and Hard of Hearing (SHH). SignWiki is part of the Finnish Sign Language Corpus and Citizen Dictionary Project, that ran from 2013–15. Since the end of the project the Finnish Association of the Deaf is responsible for maintaining the sites.

Table 5.15: Fact Sheet: Finland-Swedish SignWiki

Name	Finland-Swedish SignWiki
Language	FinSSL, Finnish
Size	5,704 signs
Linguistic information	Citation form, pictures, lexical category, phonological information
Licence	<i>not available</i>
Access	Public access via browsable homepage
Webpage	https://finssl.signwiki.org
Institution	Finnish Association of the Deaf
Publications	<i>not available</i>

5.16 FINNISH SIGNBANK

Finnish Signbank is a lexical database of FinSL and FinSSL that is linked to the Corpus of Finnish Sign Language ([Section 3.5](#)). It is developed by the Corpus Project of Finland's Sign Languages (CFINSL) at the Sign Language Centre of the University of Jyväskylä in cooperation with the corpus and dictionary work of the Finnish Association of the Deaf.

On the basis of the Finnish Signbank two lexica are being created: the Kipo Corpus lexicon of the Finnish Association of the Deaf and a lexicon from the CFINSL project.

The Finnish Signbank runs on the software FinSL-signbank⁷, a fork of NGT Signbank, which was in turn based on Auslan Signbank (cf. introduction to [Chapter 5](#) and [Section 5.21](#)).

Table 5.16: Fact Sheet: Finnish Signbank

Name	Finnish Signbank
Language	FinSL, FinSSL, Finnish
Size	3,000 signs
Linguistic information	Citation form, gloss, information about source, Finnish translations, relations between glosses
Licence	CC BY-NC-SA 4.0
Access	Public access via browsable homepage
Webpage	https://signbank.csc.fi/
Institution	University of Jyväskylä
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/finnishsignbank.html

5.17 FINNISH SIGNWIKI

The Finnish SignWiki is a multifaceted open access dictionary of FinSL. It used crowdsourcing to collect the displayed information. Signs and other information can be uploaded and discussed online. A related SignWiki for FinSSL also exists (see [Section 5.15](#)). SignWiki was created in a cooperation between the Finnish Association of the Deaf, the Humanities University of Applied Sciences (Humak) and the Icelandic Communication Center for the Deaf and Hard of Hearing (SHH). SignWiki is part of the Finnish Sign Language Corpus and Citizen Dictionary Project, that ran from 2013–15. Since the end of the project the Finnish Association of the Deaf is responsible for maintaining the sites.

Table 5.17: Fact Sheet: Finnish SignWiki

Name	Finnish SignWiki
Language	FinSL, Finnish
Size	2,629 signs
Linguistic information	Citation form, pictures, lexical category, phonological information
Licence	<i>not available</i>
Access	Public access via browsable homepage
Webpage	https://finsl.signwiki.org

⁷<https://github.com/Signbank/FinSL-signbank>

Table 5.17: Fact Sheet: Finnish SignWiki (cont.)

Name	Finnish SignWiki
Institution	Finnish Association of the Deaf
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/finnishsignwiki.html

5.18 GALEX

GaLex is a corpus-based dictionary of DGS for technical terms from the field of landscaping and horticulture. It was built at the Institute of German Sign Language (IDGS), University Hamburg from 2006–2009. The project was led by Siegmund Prillwitz and Christian Rathmann.

Table 5.18: Fact Sheet: GaLex

Name	GaLex
Language	DGS, German, English
Size	710 signs
Linguistic information	Citation form, German and English translations, definition, semantic grouping, HamNoSys transcription
Licence	<i>not available</i>
Access	Public access via browsable homepage
Webpage	https://www.sign-lang.uni-hamburg.de/galex
Institution	Universität Hamburg
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/galex.html

5.19 GLEX

GLex is a corpus-based dictionary of DGS for technical terms from the field of health and nursing care. It was built at the Institute of German Sign Language (IDGS), University Hamburg from 2004–2007. The project was led by Siegmund Prillwitz and Reiner Konrad.

Table 5.19: Fact Sheet: GLex

Name	GLex
Language	DGS, German, English
Size	2,330 signs
Linguistic information	Citation form, German and English translations, definition, semantic grouping, HamNoSys transcription
Licence	<i>not available</i>
Access	Public access via browsable homepage

Table 5.19: Fact Sheet: GLex (cont.)

Name	GLex
Webpage	https://www.sign-lang.uni-hamburg.de/glex
Institution	Universität Hamburg
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/glex.html

5.20 GLOBAL SIGNBANK - LSFb

The dataset of LSFb at the Global Signbank is a restricted dataset of LSFb signs collected in Namur and enriched in the VICI project by Onno Crasborn.

Table 5.20: Fact Sheet: Global Signbank - LSFb

Name	Global Signbank - LSFb
Language	LSFB, French, English
Size	3,512 signs
Linguistic information	<i>not available</i>
Licence	<i>not available</i>
Access	Restricted access
Webpage	https://signbank.science.ru.nl/datasets/LSFB
Institution	<i>not available</i>
Publications	<i>not available</i>

5.21 GLOBAL SIGNBANK - NGT

The dataset of NGT in the Global Signbank was built on data from the ECHO Corpus (Section 3.13), Corpus NGT (Section 3.4) and Handy Signs. The lexicon data is directly linked to the Corpus NGT. The database was built at Radboud University under the lead of Onno Crasborn.

The NGT dataset is connected to *Concepticon*⁸ (List et al., 2021), a resource connecting different concept lists used in linguistic literature (see Börstell et al., 2020).

The software of Global Signbank⁹ was originally developed as an NGT Signbank fork of the original Auslan Signbank software that added features relevant to NGT research. It was then further extended to support multiple languages, resulting in Global Signbank (cf. introduction to Chapter 5).

⁸<https://concepticon.clld.org/>

⁹<https://github.com/Signbank/Global-signbank>

Table 5.21: Fact Sheet: Global Signbank - NGT

Name	Global Signbank - NGT
Language	NGT, Dutch, English
Size	4,167 signs
Linguistic information	Glosses in Dutch and English, citation forms, translation into Dutch and English, disambiguation word combinations in Dutch, phonetic information
Licence	CC BY-NC-SA 4.0
Access	3,950 signs public accessible via homepage, confirmed registered user can access more data
Webpage	https://signbank.science.ru.nl/datasets/NGT
Institution	Radboud University
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/globalsignbankngt.html

5.22 GLOBAL SIGNBANK - NTS

The dataset of NTS at the Global Signbank is a restricted dataset of Norwegian Sign Language (NTS) signs collected by Lindsay Ferrara at the Norwegian University of Science and Technology.

Table 5.22: Fact Sheet: Global Signbank - NTS

Name	Global Signbank - NTS
Language	NTS, Norwegian, English
Size	1,260 signs
Linguistic information	<i>not available</i>
Licence	<i>not available</i>
Access	Restricted access
Webpage	https://signbank.science.ru.nl/datasets/NTS
Institution	Norwegian University of Science and Technology
Publications	<i>not available</i>

5.23 GLOBAL SIGNBANK - VGT

The dataset of VGT at the Global Signbank is a restricted dataset of VGT signs collected by Onno Crasborn and Sam Verstraete. The dataset is owned by the Vlaams Gebarentaal Centrum, Antwerp.

Members of Radboud University and Vlaams Gebarentaal Centrum are allowed to use the dataset.

Table 5.23: Fact Sheet: Global Signbank - VGT

Name	Global Signbank - VGT
Language	VGT, Dutch, English
Size	16,928 signs
Linguistic information	<i>not available</i>
Licence	<i>not available</i>
Access	Restricted access
Webpage	https://signbank.science.ru.nl/datasets/VGT
Institution	Vlaams Gebarentaal Centrum, Antwerp
Publications	<i>not available</i>

5.24 HLEX

HLex is a corpus-based dictionary of DGS for technical terms from the field of home economics. It was built at the Institute of German Sign Language (IDGS), University Hamburg from 1998–2000. The project was led by Siegmund Prillwitz.

Table 5.24: Fact Sheet: HLex

Name	HLex
Language	DGS, German
Size	1,560 signs
Linguistic information	Citation form, German and English translations, definition, semantic grouping, HamNoSys transcription
Licence	<i>not available</i>
Access	Temporarily unavailable at the time of writing
Webpage	https://www.sign-lang.uni-hamburg.de/hlex
Institution	Universität Hamburg
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/hlex.html

5.25 LEDASILA

LedaSila is a lexical database for SLs. At the time of writing it containing only signs for Austrian Sign Language (ÖGS). The ÖGS dataset collects signs from different projects. The database was built at the Fakultätszentrum für Gebärdensprache und Hörbehindertenkommunikation (ZGH) at the Alpen-Adria-Universität Klagenfurt. Signs are searchable via IDs and German keywords.

Table 5.25: Fact Sheet: LedaSila

Name	LedaSila
Language	ÖGS, German
Size	17,400 signs
Linguistic information	Category, region, translations, citation forms, number of hands, mouthing
Licence	Creative Commons (unspecified)
Access	Public access via browsable homepage
Webpage	https://ledasila.aau.at
Institution	Alpe-Adria-Universität Klagenfurt
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/ledasila.html

5.26 LEX-LSFB

Lex-LSFB is a lexical database developed on the basis of the annotation within the Corpus LSFB project. All fully-lexical signs identified in the annotated videos have been collected and made available. Lex-LSFB is built at the French Belgian Sign Language Laboratory (LSFB-Lab) under the lead of Laurence Meurant.

Lex-LSFB is connected to the Corpus LSFB (Section 3.3) so that sign glosses and French keywords found in the lexical database can be used to search for signs in videos. As corpus annotation still continues Lex-LSFB will be regularly expanded.

Table 5.26: Fact Sheet: Lex-LSFB

Name	Lex-LSFB
Language	LSFB, French
Size	<i>not available</i>
Linguistic information	ID-gloss, french translation(s), citation form, link to LSFB en ligne if available
Licence	CC BY-NC-SA 4.0
Access	Public access via browsable homepage
Webpage	https://www.corpus-lsfb.be/lexique.php
Institution	French Belgian Sign Language Laboratory (LSFB-Lab)
Publications	<i>not available</i>

5.27 LSE-SIGN

LSE-Sign is a lexical database created as a research tool that contains signs and non-signs. The database was developed by the Basque Center on Cognition, Brain and Language (BCBL)

in collaboration with Fundación CNSE (The State Confederation for Deaf People of Spain).

LSE-Signs contains signs from the first standardised LSE dictionary published by the Spanish National Association of Deaf People and non-signs that were generated by altering one of the principal phonological parameters of a real sign. Signs are shown without mouthing or emotional content and coded according to formal and grammatical criteria. Signs and non-signs can be searched by their properties and display the results with an adjustable amount of information.

Table 5.27: Fact Sheet: LSE-Sign

Name	LSE-Sign
Language	LSE, Spanish
Size	2,400 signs and 2,700 non-signs
Linguistic information	Coding for grammatical, phonological, articulatory information (handshape, location, movement, non-manual elements)
Licence	LSE-Sign license http://lse-sign.bcbl.eu/web-busqueda/?page_id=8
Access	Restricted access requires confirmed registration
Webpage	http://lse-sign.bcbl.eu/web-busqueda/
Institution	Basque Center on Cognition, Brain and Language (BCBL), Fundación CNSE
Publications	Gutierrez-Sigut et al. (2016)

5.28 LSFb EN LIGNE

Dictionnaire de LSFb en ligne is an online dictionary of LSFb built on the basis of Corpus LSFb (Section 3.3) data collected by LSFb asbl, where the project is based. Signs are searchable by handshape, French keywords and topics.

Table 5.28: Fact Sheet: LSFb en ligne

Name	LSFB en ligne
Language	LSFB, French
Size	4,000 signs
Linguistic information	Citation form, homonyms, synonyms, variants, SignWriting, etymology, definition, signed example, translations to French
Licence	<i>not available</i>
Access	Public access via browsable homepage
Webpage	http://dicto.lsfb.be/
Institution	LSFB asbl
Publications	<i>not available</i>

5.29 NOEMA

NOEMA is an electronic dictionary of GSL/Greek for a large non-specialised audience. NOEMA was created at the Institute for Language and Speech Processing at the Athena Research Center. Signs are searchable via handshapes and combinations of handshapes and are categorised into thematic groups.

Table 5.29: Fact Sheet: NOEMA

Name	NOEMA
Language	GSL, Greek
Size	3,000 signs
Linguistic information	Lemma, semantic encoding, antonyms, synonyms, use, notes, HamNoSys transcript, translation to Greek
Licence	<i>not available</i>
Access	Restricted access requires individual license agreement
Webpage	Meta share entry: http://metashare.ilsp.gr:8080/repository/browse/basic-vocabulary-of-the-greek-sign-language/2cd9b952609f11e2918d842b2b6a04d7b614b9b36fad487582ca52a4eb0a6473/
Institution	Institute for Language and Speech Processing, Athena Research Center
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/noema.html

5.30 OCELLES

Ocelles is a collaborative website entirely bilingual in French and LSF which collects signs, definitions, information on projects and organisations. For each concept at least one definition and its associated descriptors in various knowledge fields are proposed. Users can upload information (e. g. texts, pictures, videos, presentation) which is examined by experts on form and content before being released online.

Table 5.30: Fact Sheet: Ocelles

Name	Ocelles
Language	LSF, French
Size	<i>not available</i>
Linguistic information	Citation form, definitions, example sentences, related definitions, explanations, stories
Licence	CC BY-SA 2.0 FR
Access	Public access via browsable homepage
Webpage	http://www.ocelles.fr

Table 5.30: Fact Sheet: Ocelles (cont.)

Name	Ocelles
Institution	Ocelles
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/ocelles.html

5.31 PLEX

PLex is a corpus-based dictionary of DGS for technical terms from the field of psychology. It was built at the Institute of German Sign Language (IDGS), University Hamburg from 1993–1995. The project was led by Siegmund Prillwitz.

Table 5.31: Fact Sheet: PLEX

Name	PLex
Language	DGS, German
Size	1,270 signs
Linguistic information	Citation form, German translations, definition, semantic grouping, HamNoSys transcription
Licence	<i>not available</i>
Access	Temporarily unavailable at the time of writing
Webpage	https://www.sign-lang.uni-hamburg.de/projekte/plex
Institution	Universität Hamburg
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/plex.html

5.32 POLYTROPON

The POLYTROPON lexicon is a collection of existing lexical resources of GSL which were enriched with new lemmas. Main resource for the lemma list were the NOEMA dictionary (Section 5.29) and the GSL dataset of the Dicta-Sign corpus (Section 3.10). The POLYTROPON corpus was constructed at the Athena Research Center at the Institute for Language and Speech Processing (ILSP) under the lead of Eleni Efthimiou.

A by-product of the lexical resource is the POLYTROPON parallel corpus (Section 3.20), which contains GSL examples of use for every sign entry within the POLYTROPON lexicon.

Table 5.32: Fact Sheet: POLYTROPON

Name	POLYTROPON
Language	GSL, Greek
Size	8,616 signs

Table 5.32: Fact Sheet: POLYTROPON (cont.)

Name	POLYTROPON
Linguistic information	Citation form, usage examples, exhaustive coding of lemmas for manual and non-manual features, ontology scheme
Licence	Individual license agreement may be possible
Access	Not available online at the time of writing
Webpage	Temporarily unavailable at the time of writing
Institution	Institute for Language and Speech Processing, Athena Research Center
Publications	<i>not available</i>

5.33 SEMATOS

sematos is a collection of five dictionaries: four for European SLs (Catalan Sign Language (LSC), LSE, , LSF, DGS) and one for IS. Signs are searchable by handshape, topic, grammatical categorisation and translation equivalents. The homepage also offers games and a message board for exchange.

Table 5.33: Fact Sheet: sematos

Name	sematos
Language	LSC, LSE, LSF, IS, DGS, Spanish, French, English, German
Size	297 LSC signs, 6,076 LSE signs, 3,605 LSF signs, 440 IS signs, unknown number of DGS signs
Linguistic information	Keywords, handshape, POS
Licence	<i>not available</i>
Access	Public access via browsable homepage; DGS dataset temporarily unavailable at the time of writing
Webpage	http://www.sematos.eu
Institution	sematos
Publications	<i>not available</i>

5.34 SGB-FSS LEXICON

The SGB-FSS lexicon is a multilingual lexicon of DSGS/German, LSF/French and LIS/Italian. The lexicon is built by the SGB-FSS – the Swiss Association of the Deaf. Signs are searchable by keywords in German, French and Italian.

Table 5.34: *Fact Sheet: SGB-FSS Lexicon*

Name	SGB-FSS Lexicon
Language	DSGS, LIS, LSF, German, Italian, French
Size	<i>not available</i>
Linguistic information	Gloss, description, definition, category, example sentence, illustration, photograph
Licence	Restricted access requires individual license agreement
Access	Public access via browsable homepage
Webpage	https://signsuisse.sgb-fss.ch/de/
Institution	SGB-FSS Schweizer Gehörlosenbund (Swiss Association of the Deaf)
Publications	<i>not available</i>

5.35 SIGNSTUDY

SignStudy is an online lexical resource for ASL signs. The signs are structured in terms of semantic categories and subcategories and annotated by 67 handshapes, 38 semantic categories, and 238 semantic subcategories. SignStudy was created by SignSchool, an online ASL learning platform.

Signs are searchable by English keywords. SignStudy – as well as ASL Signbank (Section 5.1) and ASL-Lex (Section 5.2) – is used in the construction of *ASLNet* (C. P. Lualdi et al., 2021), a wordnet for ASL.

Table 5.35: *Fact Sheet: SignStudy*

Name	SignStudy
Language	ASL, English
Size	4,500 signs
Linguistic information	Citation form, synonyms, polysemous words, parameters, semantic categories and semantic hierarchies on two levels
Licence	<i>not available</i>
Access	Temporarily unavailable at the time of writing
Webpage	http://www.signstudy.org/
Institution	SignSchool
Publications	C. Lualdi et al. (2019)

5.36 SLEX

SLex is a corpus-based dictionary of DGS for technical terms from the field of social work and social pedagogics. It was built at the Institute of German Sign Language (IDGS), University Hamburg from 2001–2003. The project was led by Siegmund Prillwitz.

Table 5.36: Fact Sheet: SLex

Name	SLex
Language	DGS, German, English
Size	940 signs
Linguistic information	Citation form, German and English translations, definition, semantic grouping, HamNoSys transcription
Licence	<i>not available</i>
Access	Temporarily unavailable at the time of writing
Webpage	https://www.sign-lang.uni-hamburg.de/slex
Institution	Universität Hamburg
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/slex.html

5.37 SLOVAR SZJ

Slovar SZJ is a dictionary for SZJ/Slovenian from the Slovene Association of the Deaf. Signs are searchable by Slovenian keywords.

Table 5.37: Fact Sheet: Slovar SZJ

Name	Slovar SZJ
Language	SZJ, Slovenian
Size	<i>not available</i>
Linguistic information	Citation form, definition, children stories
Licence	<i>not available</i>
Access	Public access via browsable homepage
Webpage	https://www.szj.si/
Institution	Zveza društev gluhih in naglušnih Slovenije (Slovene Association of the Deaf)
Publications	<i>not available</i>

5.38 SSLD

The Swedish Sign Language Dictionary is an online dictionary for STS. The dictionary was created at the Department of Linguistics of Stockholm University.

SSLD was first built as an online video representation of a printed dictionary (Hedberg, 1998) which was extended over the years. The Swedish Sign Language Corpus (Section 3.24) as well as the STS community are used to collect and document signs and lexical variation for the SSLD.

Table 5.38: Fact Sheet: SSLD

Name	SSLD
Language	STS, Swedish
Size	19,149 signs
Linguistic information	Citation form, ID-number, description, example sentences, picture, translations to Swedish and English, meanings, alternative signs, handshape, topic, corpus evidence, homonyms, synonyms, phonological description and transcription, glosses
Licence	CC BY-NC-SA 2.5 SE
Access	Public access via browsable homepage
Webpage	https://teckensprakslexikon.su.se/
Institution	University Stockholm
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/ssld.html

5.39 SUVI OR BASIC DICTIONARY OF FINSL

Suvi is an online dictionary of FinSL. The dictionary contains signs in the subject areas of basic signs, numbers and expressions of numbers, sexuality and babies, place names, fairy tales and health care signs. Suvi was built in 2003 in a collaboration of the Finnish Association of the Deaf, the Finnish Language Research Center and the Blue Meteorite as well as Microsoft in Finland. Suvi is regularly expanded and improved. Today the Finnish Association of the Deaf is responsible for maintaining the site.

Signs are searchable by the sign structure and by Finnish keywords.

Table 5.39: Fact Sheet: SUVI or Basic Dictionary of FinSL

Name	SUVI or Basic Dictionary of FinSL
Language	FinSL, Finnish
Size	1,211 signs
Linguistic information	Citation form, handshape, number of hands, position, movement, variants, plural, example sentences
Licence	<i>not available</i>
Access	Public access via browsable homepage
Webpage	http://suvi.viittomat.net
Institution	Finnish Association of the Deaf
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/suvi.html

5.40 SWISS GERMAN SIGN LANGUAGE LEXICON

The Lexicon of Swiss German Sign Language is arranged into different sub-lexicons. Two of them, containing 685 technical terms in the domains of nutrition and economy, are described in more detail. The lexicon was built at the University of Zurich under the lead of Penny Boyes-Braem.

Signs can be searched by German keywords, by domain restrictions, by a given status (used, known, new) or by glosses. To automatically obtain candidates for semantic relations in Swiss-German Sign Language (DSGS) the German-language wordnet *Germanet*¹⁰ (Hamp and Feldweg, 1997) was linked to the DSGS lexicon (see Ebling et al., 2012).

Table 5.40: *Fact Sheet: Swiss German Sign Language Lexicon*

Name	Swiss German Sign Language Lexicon
Language	DSGS, German
Size	9,000 signs
Linguistic information	For 3,000 signs: citation form, meaning, morphological and syntactic characteristics, usage, drawings, HamNoSys and SignWriting transcription, German translations
Licence	Individual license agreement may be possible
Access	Temporarily unavailable at the time of writing
Webpage	https://linguistik-signlang.uzh.ch
Institution	University of Zurich
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/dsgslexicon.html

5.41 TLEX

TLex is a corpus-based dictionary of DGS for technical terms from the field of joinery. It was built at the Institute of German Sign Language (IDGS), University Hamburg from 1996–1998. The project was led by Siegmund Prillwitz.

Table 5.41: *Fact Sheet: TLex*

Name	TLex
Language	DGS, German
Size	2,800 signs
Linguistic information	Citation form, German and English translations, definition, semantic grouping, HamNoSys transcription
Licence	<i>not available</i>
Access	Temporarily unavailable at the time of writing

¹⁰<https://uni-tuebingen.de/en/142806>

Table 5.41: Fact Sheet: TLex (cont.)

Name	TLex
Webpage	https://www.sign-lang.uni-hamburg.de/tlex
Institution	Universität Hamburg
Publications	sign-lang@LREC Anthology: https://www.sign-lang.uni-hamburg.de/lrec/data/tlex.html

6 CONCLUSIONS

The overview of corpora shows that at least the publicly available parts of most corpora have a rather basic annotation layer, covering translation and lemmatisation, but mostly no more supra-segmental annotation, i. e. grammatical structures cannot directly be identified in the data, but need to be inferred.

While this is most understandable given the huge amount of work to be invested in producing such resources and making them available to the research community, the research field would profit substantially from additional data available internally in the data creator projects or created by other researchers even if they cover only selected parts of the corpus data. As a first step, it might be helpful to the research community if metadata formats were available that allowed the specification of available annotation layers independent of whether they are (yet) publicly accessible. That way, researchers interested in specific phenomena would at least know whom to contact instead of starting from scratch.

Within EASIER one of the following steps will be the harmonisation of available annotated data. Some first commonalities within annotation conventions are becoming visible in the Overview and can be used as a ground for the harmonisation process, e. g. the annotation conventions of the Auslan corpus ([Johnston, 2010](#)) are the basis for conventions of many other datasets. Awareness of such commonalities will be important when working towards an annotation standard for EASIER.

The overview of elicitation tasks across corpora might be helpful when looking for specific language patterns. As an example, process descriptions might be a good starting point when looking for ground-figure constructions.

It is our hope that this document will be useful both for the efforts within EASIER as well as presenting a useful source of information for the SL research community in general.

ACKNOWLEDGEMENTS

We would like to thank all researchers who contributed to this report by providing information on their corpora, giving us access to their data or bringing information on their data up to date.

REFERENCES

- Börstell, Carl, Onno Crasborn, and Lori Whynot (2020). “Measuring Lexical Similarity across Sign Languages in Global Signbank”. 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. 21–26. ISBN: 979-10-95546-54-2. URL: <https://www.sign-lang.uni-hamburg.de/lrec/pub/20011.pdf>.
- Brien, David (1992). *Dictionary of British Sign Language/English*. In collab. with University of Durham, Deaf Studies Research Unit, and British Deaf Association. London; Boston: Faber and Faber. 1084 pp. ISBN: 978-0-571-14346-7.
- Briggs, Raymond (1978). *The Snowman*. London: Hamish Hamilton. 32 pp. ISBN: 978-0-394-83973-8.
- Chafe, Wallace L., ed. (1980). *The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production*. Advances in discourse processes 3. Norwood, New Jersey, USA: Ablex Publishing Corporation. 327 pp. ISBN: 978-0-89391-032-7.
- Cormier, Kearsy, Jordan Fenlon, Sannah Gulamani, and Sandra Smith (2017). *BSL Corpus Annotation Conventions*. Annotation Convention. Version 3.0. London, United Kingdom: Deafness Cognition and Language (DCAL) Research Center, University College London, p. 19. URL: https://bslcorpusproject.org/wp-content/uploads/BSLCorpus_AnnotationConventions_v3.0_March2017.pdf (visited on 07/21/2021).
- Crasborn, Onno (2010). “What does “Informed Consent” Mean in the Internet Age? Publishing Sign Language Corpora as Open Content”. In: *Sign Language Studies* 10.2, pp. 276–290. ISSN: 0302-1475. DOI: [10.1353/sls.0.0044](https://doi.org/10.1353/sls.0.0044). JSTOR: [26190572](https://www.jstor.org/stable/26190572).
- Crasborn, Onno and Thomas Hanke (2003). *Additions to the IMDI metadata set for sign language corpora. Agreements at an ECHO workshop*. Workshop Report. Version 12.01.2010. Nijmegen, Netherlands: Radboud University, p. 7. URL: http://www.sign-lang.ruhosting.nl/echo/docs/SignMetadata_Oct2003.pdf (visited on 10/30/2020).
- Crasborn, Onno and Han Sloetjes (2008). “Enhanced ELAN functionality for sign language corpora”. 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 Zwitterlood. Paris, France: European Language Resources Association (ELRA), pp. 39–43. URL: <https://www.sign-lang.uni-hamburg.de/lrec/pub/08022.pdf>.
- Crasborn, Onno, Inge Zwitterlood, Els van der Kooij, and Richard Bank (2020). *Annotation conventions for the Corpus NGT*. Annotation Convention. Version 4. Nijmegen, Netherlands: Centre for Language Studies & Department of Linguistics, Radboud University. URL: https://www.ru.nl/publish/pages/1013556/corpusngt_annotationconventions_v4_1.pdf (visited on 07/22/2021).
- Davies, John Howard, John Birkin, and Paul Weiland, director (1990–1995). *Mr. Bean*. United Kingdom: ITV: Thames Television (1990–1992), Central Independent Television (1993–1995).

- 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>.
- Fišer, Darja and Benoît Sagot (2015). “Constructing a poor man’s wordnet in a resource-rich world”. In: *Language Resources and Evaluation* 49.3, pp. 601–635. ISSN: 1574-0218. DOI: [10.1007/s10579-015-9295-6](https://doi.org/10.1007/s10579-015-9295-6).
- Freleng, Friz, director (1950). *Canary Row*. scriptwriter Tedd Pierce, Mel Blanc, and Bea Benaderet. Merrie Melodies #1136. Warner Bros.
- Gutierrez-Sigut, Eva, Brendan Costello, Cristina Baus, and Manuel Carreiras (2016). “LSE-Sign: A lexical database for Spanish Sign Language”. In: *Behavior Research Methods* 48.1, pp. 123–137. ISSN: 1554-3528. DOI: [10.3758/s13428-014-0560-1](https://doi.org/10.3758/s13428-014-0560-1).
- 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 (2002). “iLex - A Tool for Sign Language Lexicography and Corpus Analysis”. In: *3rd International Conference on Language Resources and Evaluation (LREC 2002)* (Las Palmas, Canary Islands, Spain). Ed. by Manuel González Rodríguez and Carmen Paz Suarez Araujo. Paris, France: European Language Resources Association (ELRA), pp. 923–926. URL: <http://www.lrec-conf.org/proceedings/lrec2002/pdf/330.pdf>.
- (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, Elena Jahn, Sabrina Wähl, Oliver Böse, and Lutz König (2020). “SignHunter – A Sign Elicitation Tool Suitable for Deaf Events”. 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. 83–88. ISBN: 979-10-95546-54-2. URL: <https://www.sign-lang.uni-hamburg.de/lrec/pub/20030.pdf>.
- Hanke, Thomas and Jakob Storz (2008). “iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography”. 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 Zwitterlood. Paris, France: Eu-

- European Language Resources Association (ELRA), pp. 64–67. URL: <https://www.sign-lang.uni-hamburg.de/lrec/pub/08011.pdf>.
- Hedberg, Thomas, ed. (1998). *Svenskt teckenspråkslexikon*. In collab. with Mikael Lindström and Syed Ahmed Mumtaz. Trans. by Tommy Lyxell and Elisabet Tersmeden-Lundquist. Leksand, Sweden: Sveriges dövas riksförbund (SDR). ISBN: 91-630-5542-2.
- Heßmann, Jens (2001). *Gehörlos so! Materialien zur Gebärdensprache*. 2 vols. International studies on sign language and the communication of the deaf 32. Hamburg, Germany: Signum Verlag. 1032 pp. ISBN: 3-927731-86-2.
- Hughes, Patrick, director (2008). *Signs*. scriptwriter Patrick Hughes, Karl Fleet, Nick Worthington, Victoria Conners Bell, Julian Dimsey, and Jodie Hari. Short film. IndieFlix.
- Johnston, Trevor (2008). “Corpus linguistics and signed languages: no lemmata, no corpus”. 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 Zwitterlood. Paris, France: European Language Resources Association (ELRA), pp. 82–87. URL: <https://www.sign-lang.uni-hamburg.de/lrec/pub/08031.pdf>.
- (2010). “From archive to corpus: Transcription and annotation in the creation of signed language corpora”. In: *International Journal of Corpus Linguistics* 15.1, pp. 106–131. ISSN: 1569-9811. DOI: [10.1075/ijcl.15.1.05joh](https://doi.org/10.1075/ijcl.15.1.05joh).
- Johnston, Trevor and Adam Schembri (2006). “Issues in the creation of a digital archive of a signed language”. In: *Sustainable Data from Digital Fieldwork: From creation to archive and back*. Sidney, Australia: Sydney University Press, pp. 7–16. ISBN: 978-1-920898-50-2. URL: <http://hdl.handle.net/2123/1289>.
- Kanis, Jakub, Jiří Zahradil, Filip Jurčíček, and Luděk Müller (2006). “Czech-Sign Speech Corpus for Semantic Based Machine Translation”. In: *9th International Conference on Text, Speech and Dialogue*. TSD 2006. Ed. by Petr Sojka, Ivan Kopeček, and Karel Pala. Lecture Notes in Computer Science. Brno, Czech Republic: Springer, pp. 613–620. ISBN: 978-3-540-39091-6. DOI: [10.1007/11846406_77](https://doi.org/10.1007/11846406_77).
- Konrad, Reiner, Thomas Hanke, Gabriele Langer, Susanne König, Lutz König, Rie Nishio, and Anja Regen (2020). *Öffentliches DGS-Korpus: Annotationskonventionen / Public DGS Corpus: Annotation Conventions*. Project Note AP03-2018-01. Version 3. Hamburg, Germany: DGS-Korpus project, IDGS, Hamburg University. DOI (latest version): [10.25592/uh-hfdm.822](https://doi.org/10.25592/uh-hfdm.822).
- List, Johann Mattis, Christoph Rzymyski, Simon Greenhill, Nathanael Schweikhard, Kristina Pi- anykh, Annika Tjuka, Carolin Hundt, and Robert Forkel (2021). *CLLD Concepticon 2.5.0*. Version 2.5.0. Max Planck Institute for Evolutionary Anthropology. DOI: [10.5281/zenodo.4911605](https://doi.org/10.5281/zenodo.4911605).
- Lualdi, Colin, Jack Hudson, Christiane Fellbaum, and Noah Buchholz (2019). “Building ASLNet, a Wordnet for American Sign Language”. In: *Proceedings of the 10th Global Wordnet Conference*. GWC 2019. Wrocław, Poland: Global Wordnet Association, pp. 315–322. ISBN: 978-83-7493-108-3. URL: <https://aclanthology.org/2019.gwc-1.40>.
- 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*. GWC 2021. 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).
- Lucas, Ceil, Robert Bayley, and Clayton Valli (2002). *Sociolinguistic Variation in American Sign Language*. In collab. with Mary Rose, Alyssa Wulf, Paul Dudis, Susan Schatz, and Laura

- Sanheim. The sociolinguistics in deaf communities series 7. Washington, DC: Gallaudet Univ. Press. 237 pp. ISBN: 978-1-56368-113-4.
- Matthes, Silke, Thomas Hanke, Jakob Storz, Eleni Efthimiou, Athanasia-Lida Dimou, Panagiotis Karioris, Annelies Braffort, Annick Choisier, Julia Pelhate, and Eva Safar (2010). "Elicitation tasks and materials designed for Dicta-Sign's multi-lingual corpus". 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. 158–163. URL: <https://www.sign-lang.uni-hamburg.de/lrec/pub/10019.pdf>.
- Mayer, Mercer (1969). *Frog, Where Are You?* New York: Dial Books for Young Readers. 32 pp. ISBN: 978-0-8037-2881-3.
- Nishio, Rie, Sung-Eun Hong, Susanne König, Reiner Konrad, Gabriele Langer, Thomas Hanke, and Christian Rathmann (2010). "Elicitation methods in the DGS (German Sign Language) Corpus Project". 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. 178–185. URL: <https://www.sign-lang.uni-hamburg.de/lrec/pub/10026.pdf>.
- Nonhebel, Annika, Onno Crasborn, and Els van der Kooij (2004). *Sign language transcription conventions for the ECHO Project*. Annotation Convention. Version 9. ECHO project, University of Nijmegen. URL: <http://hdl.handle.net/2066/57889> (visited on 07/21/2021).
- 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.
- Salonen, Juhana, Tuija Wainio, Antti Kronqvist, and Jarkko Keränen (2019). *Suomen Viittomakielten Korpusprojektin (CFINSL) Annotointiohjeet*. Annotation Convention. Helmikuu: Department of Linguistics and Communication Sciences, Sign Language Center, University of Jyväskylä, p. 40. URL: https://www.jyu.fi/hytk/fi/laitokset/kivi/opiskelu/tutkinto-ohjelmat-ja-oppiaineet/viittomakieli/tutkimus-2/suomen-viittomakielten-korpusprojekti/cfinsl_annotointiohjeet_2019_2versio.pdf (visited on 07/28/2021).
- Volterra, Virginia, Alessandro Laudanna, Serena Corazza, Elena Radutzky, and Francesco Natale (1984). "Italian Sign Language: The Order of Elements in the Declarative Sentence". In: *Recent research on European sign languages*. Ed. by Filip Loncke, Penny Boyes-Braem, and Yvan Lebrun. Lisse, Netherlands: Swets & Zeitlinger, pp. 19–48. ISBN: 978-90-265-0465-5.
- Wahl, Sabrina, Gabriele Langer, and Anke Müller (2018). "Hand in Hand - Using Data from an Online Survey System to Support Lexicographic Work". 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. 199–206. ISBN: 979-10-95546-01-6. URL: <https://www.sign-lang.uni-hamburg.de/lrec/pub/18025.pdf>.

Wallin, Lars and Johanna Mesch (2018). *Annoteringskonventioner för teckenspråkstexter*. Annotation Convention. Version 7. Stockholm, Sweden: Avdelningen för teckenspråk, Stockholms universitet, p. 49. URL: <http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-152163> (visited on 07/21/2021).

