EUROPEAN LANGUAGE EQUALITY

D1.40

Report on Europe's Sign Languages

Authors	Vincent Vandeghinste, Mirella De Sisto, Maria Kopf, Marc Schulder, Caro Brosens, Lien Soetemans, Rehana Omardeen, Frankie Picron, Davy Van Landuyt, Irene Murtagh, Eleftherios Avramidis, Mathieu De Coster
Dissemination level	Public
Date	09-06-2023

About this document

Project	European Language Equality (ELE)
Grant agreement no.	LC-01641480 – 101018166 ELE
Coordinator	Prof. Dr. Andy Way (DCU)
Co-coordinator	Prof. Dr. Georg Rehm (DFKI)
Start date, duration	01-01-2021, 18 months
Deliverable number	D1.40
Deliverable title	Report on Europe's Sign Languages
Type	Report
Number of pages	31
Status and version	Final (*Note: this document is not a contractual ELE deliverable.*)
Dissemination level	Public
Date of delivery	09-06-2023
Work package	WP1: European Language Equality – Status Quo in 2020/2021
Task	Task 1.3 Language Technology Support of Europe's Languages in 2020/2021
Authors	Vincent Vandeghinste, Mirella De Sisto, Maria Kopf, Marc Schulder, Caro Brosens, Lien Soetemans, Rehana Omardeen, Frankie Picron, Davy Van Landuyt, Irene Murtagh, Eleftherios Avramidis, Mathieu De Coster
Reviewers	Andy Way (DCU), Georg Rehm (DFKI)
Editors	Maria Giagkou, Stelios Piperidis, Georg Rehm, Jane Dunne
EC project officers	Susan Fraser, Miklos Druskoczi
Contact	European Language Equality (ELE) ADAPT Centre, Dublin City University Glasnevin, Dublin 9, Ireland Prof. Dr. Andy Way – andy.way@adaptcentre.ie European Language Equality (ELE) DFKI GmbH Alt-Moabit 91c, 10559 Berlin, Germany Prof. Dr. Georg Rehm – georg.rehm@dfki.de http://www.european-language-equality.eu © 2023 ELE Consortium

Consortium

1	Dublin City University (Coordinator)	DCU	IE
2	Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (Co-coordinator)	DFKI	DE
3	Univerzita Karlova (Charles University)	CUNI	CZ
4	Athina-Erevnitiko Kentro Kainotomias Stis Technologies Tis Pliroforias, Ton Epikoinonion Kai Tis Gnosis	ILSP	GR
5	Universidad Del Pais Vasco/ Euskal Herriko Unibertsitatea (University of the Basque Country)	UPV/EHU	ES
6	CROSSLANG NV	CRSLNG	BE
7	European Federation of National Institutes for Language	EFNIL	LU
8	Réseau européen pour l’égalité des langues (European Language Equality Network)	ELEN	FR
9	European Civil Society Platform for Multilingualism	ECSPM	DK
10	CLARIN ERIC – Common Language Resources and Technology Infrastructure as a European Research Infrastructure Consortium	CLARIN	NL
11	Universiteit Leiden (University of Leiden)	ULEI	NL
12	Eurescom (European Institute for Research and Strategic Studies in Telecommunications GmbH)	ERSCM	DE
13	Stichting LIBER (Association of European Research Libraries)	LIBER	NL
14	Wikimedia Deutschland (Gesellschaft zur Förderung freien Wissens e. V.)	WMD	DE
15	Tilde SIA	TILDE	LV
16	Evaluations and Language Resources Distribution Agency	ELDA	FR
17	Expert System Iberia SL	EXPSYS	ES
18	HENSOLDT Analytics GmbH	HENS	AT
19	Xcelerator Machine Translations Ltd. (KantanMT)	KNTN	IE
20	PANGEANIC-B. I. Europa SLU	PAN	ES
21	Semantic Web Company GmbH	SWC	AT
22	SIRMA AI EAD (Ontotext)	ONTO	BG
23	SAP SE	SAP	DE
24	Universität Wien (University of Vienna)	UVIE	AT
25	Universiteit Antwerpen (University of Antwerp)	UANTW	BE
26	Institute for Bulgarian Language “Prof. Lyubomir Andreychin”	IBL	BG
27	Sveučilište u Zagrebu Filozofski fakultet (Univ. of Zagreb, Faculty of Hum. and Social Sciences)	FFZG	HR
28	Københavns Universitet (University of Copenhagen)	UCPH	DK
29	Tartu Ulikool (University of Tartu)	UTART	EE
30	Helsingin Yliopisto (University of Helsinki)	UHEL	FI
31	Centre National de la Recherche Scientifique	CNRS	FR
32	Nyelvtudományi Kutatóközpont (Research Institute for Linguistics)	NYTK	HU
33	Stofnun Árna Magnússonar í íslenskum fræðum SAM (Árni Magnússon Inst. for Icelandic Studies)	SAM	IS
34	Fondazione Bruno Kessler	FBK	IT
35	Latvijas Universitātes Matemātikas un Informātikas institūts (Institute of Mathematics and Computer Science, University of Latvia)	IMCS	LV
36	Lietuvių Kalbos Institutas (Institute of the Lithuanian Language)	LKI	LT
37	Luxembourg Institute of Science and Technology	LIST	LU
38	Università ta Malta (University of Malta)	UM	MT
39	Stichting Instituut voor de Nederlandse Taal (Dutch Language Institute)	INT	NL
40	Språkrådet (Language Council of Norway)	LCNOR	NO
41	Instytut Podstaw Informatyki Polskiej Akademii Nauk (Polish Academy of Sciences)	IPIPAN	PL
42	Universidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Science)	FCULisbon	PT
43	Institutul de Cercetări Pentru Inteligență Artificială (Romanian Academy)	ICIA	RO
44	University of Cyprus, French and European Studies	UCY	CY
45	Jazykovedný ústav Ľudovíta Štúra Slovenskej akadémie vied (Slovak Academy of Sciences)	JULS	SK
46	Institut Jožef Stefan (Jozef Stefan Institute)	JSI	SI
47	Centro Nacional de Supercomputación (Barcelona Supercomputing Center)	BSC	ES
48	Kungliga Tekniska högskolan (Royal Institute of Technology)	KTH	SE
49	Universität Zürich (University of Zurich)	UZH	CH
50	University of Sheffield	USFD	UK
51	Universidad de Vigo (University of Vigo)	UVIGO	ES
52	Bangor University	BNGR	UK

Contents

List of Figures

List of Tables

List of Acronyms

AI	Artificial Intelligence
API	Application Programming Interface
BHC	Boarnsterhim Corpus
CL	Computational Linguistics
DLE	Digital Language Equality
ECRML	European Charter for Regional or Minority Languages
ELE	European Language Equality *(this project)*
ELE Programme	European Language Equality Programme *(the long-term, large-scale funding programme specified by the ELE project)*
ELG	European Language Grid (EU project, 2019-2022)
EU	European Union
FCPNM	Framework Convention for the Protection of National Minorities
HCI	Human Computer Interaction
KSF	Corpus of Spoken Frisian
LR	Language Resources/Resources
LT	Language Technology/Technologies
ML	Machine Learning
NLP	Natural Language Processing
NLU	Natural Language Understanding
NLG	Natural Language Generation
SL	Sign Language

Abstract

This report on Europe's Sign Languages is part of a series of language deliverables developed within the framework of the European Language Equality (ELE) project. The series seeks to not only delineate the current state of affairs for each European language, but to additionally identify the gaps and factors that hinder further development in research and technology. The survey presented here focuses on the condition of Language Technology (LT) with regard to Europe's Sign Languages, a set of languages often forgotten in the context of European Language Equality.

With the rise of the deep learning paradigm in artificial intelligence, sign language technologies become technologically feasible, provided that enough data is available to feed this data-hungry paradigm. It is exactly the quality and quantity of data that is the main bottleneck in development of well performing and useful technologies.

In the past, there have been several projects aimed at developing sign language technologies and methodologies that have been deemed of little value by the deaf communities. Co-creation and involvement of deaf communities throughout projects and development of technologies ensures that this does not happen again.

1. Introduction

This study is part of a series that reports on the results of an investigation of the level of support the European languages receive through technology. It is addressed to decision makers at the European and national/regional levels, language communities, journalists, etc. and it seeks to not only delineate the current state of affairs for each of the European languages covered in this series, but to additionally – and most importantly – identify the gaps and factors that hinder further development of research and technology. Identifying such weaknesses will lay the grounds for a comprehensive, evidence-based, proposal of required measures for achieving Digital Language Equality in Europe by 2030.

To this end, more than 40 research partners, experts in more than 30 European languages have conducted an enormous and exhaustive data collection procedure that provided a detailed, empirical and dynamic map of technology support for our languages.[1]

This Report on Europe's Sign Languages has been written under the initiative of the SignON project,[2] in cooperation with the EASIER project,[3] two large European research projects that focus on automatic sign language translation.

The report has been developed in the frame of the European Language Equality (ELE) project. With a large and all-encompassing consortium consisting of 52 partners covering all European countries, research and industry and all major pan-European initiatives, the ELE project develops a strategic research, innovation, and implementation agenda as well as a roadmap for achieving full digital language equality in Europe by 2030.

2. Europe's Sign Languages in the Digital Age

More than 70 million people worldwide are deaf (World Federation of the Deaf, 2023). In Europe alone, for approximately half a million of deaf and hard of hearing (DHH) people, sign languages are the main or preferred means of communication (Pasikowska-Schnass, 2018).

[1] The results of this data collection procedure have been integrated into the European Language Grid so that they can be discovered, browsed and further investigated by means of comparative visualisations across languages.
[2] https://signon-project.eu
[3] https://www.project-easier.eu

2.1. What are sign languages?

Sign languages are fully-fledged languages, each composed of their own unique lexicon and grammatical principles. These languages arise naturally among deaf communities, independent of the surrounding spoken languages.[4] They are not a derivation from nor a version of spoken languages (Vermeerbergen, 1997, 2006; Vermeerbergen et al., 2007; Baker et al., 2016). Given that sign languages arise like this around the world in different deaf communities, there is no universal sign language (much as there is no universal spoken language). In fact, in many countries multiple sign languages are recognised and used by culturally distinct groups; for example in the north of Belgium, Flemish Sign Language (*Vlaamse Gebarentaal*) is used while in the south of Belgium, French Belgian Sign Language (*Langue des signes de Belgique francophone*) is used.

While spoken languages largely make use of the oral-aural modality, sign languages exploit the visual-gestural modality. As a result, sign languages draw on their own, specific linguistic mechanisms (Meier, 2002).

Signers use visible articulators to communicate: the hands, face, torso and other parts of the body are needed for the communication production whereas the eyes (and/or hands, in the case of tactile sign languages used by deafblind people) are needed to perceive a signed message.

Signs consist of manual and non-manual parameters or building blocks. Manual parameters are the handshape, orientation, movement and location (Baker et al., 2016). Non-manual parameters are movements of the face and body e.g. mouth gestures, and/or facial expressions (Vermeerbergen, 1997; Baker et al., 2016).

Critical to linguistic expression in the visual-gestural modality is the so called 'signing space': the space in front of, next to and above the body of the signer, in which signs are produced. Sign languages organise linguistic expression within this space, and the placement and movement of signs within the signing space are integral to sign language grammars. For example, many sign languages exploit movement through space to convey verb agreement, or use space for tracking referents through discourse (Vermeerbergen, 1997).

Another linguistic particularity is the potential for simultaneous organization in sign languages, in contrast to the largely sequential information order of spoken languages. This means that when using both hands at the same time, signers can convey two different messages simultaneously, e.g. when the signer first points towards their interlocutor with the index finger of their right hand, and while holding this pointing sign, produces the sign for DRINK[5] with their left hand. When the signer raises their eyebrows during this utterance, they are asking a polar question. In this case, you know that the signer asks this question to this interlocutor ("you", first item of information) and that they would like to know whether this interlocutor would like to drink something (second item of information). Speech alone cannot relay these two distinct items of information simultaneously (Vermeerbergen, 1997; Vermeerbergen et al., 2007), but speakers can of course also use such non-verbal means for communication.

Besides the simultaneous organisation, sign languages are characterised by a lexicon that is partly "frozen" and partly "productive". A frozen or lexical sign is conventional in form (i.e. fixed parameters) and meaning (Baker et al., 2016). A productive sign can be described as a "mix 'n match sign" (Brennan, 1990), because a signer himself selects and combines the building blocks on the spot to convey a particular meaning in a specific context. Thus, the meaning of a productive sign depends on the context: the same productive sign can convey another meaning in a different context (Baker et al., 2016).

[4] Note that we use the term *spoken languages* to denote natural languages which use sound as their primary medium of communication, irrespective of whether they are considered in their auditive (speech) or written form.

[5] Capital letters are conventionally used in sign language linguistics to represent signs.

The relation between the form and meaning of a sign can be motivated or not. When the form of the sign (e.g. the handshape) refers to the concept expressed by the sign (e.g. the sign for APPLE has a handshape that denotes the round form of an apple), then this sign is called an "iconic" sign. As sign languages are produced and perceived in a visual-spatial modality, elements of visual iconicity are highly integrated into linguistic structures (Baker et al., 2016).

2.2. How sign languages differ from spoken languages

The most blatantly obvious way in which sign languages and spoken languages differ from one another is their modality, as mentioned above. Other more intangible linguistic and cultural differences stem from centuries of linguistic and societal oppression. The 'hidden' existence of sign languages and their almost non-existent roles within larger society shaped these languages and the cultures of their users (De Weerdt et al., 2003; Ladd, 2003; Vermeerbergen and Van Herreweghe, 2008; Beelaert et al., 2009).

For a long time, sign languages were not seen as full, natural languages (Vermeerbergen, 2006). Only in the 1950's did people start researching sign languages, from which grew a (renewed) appreciation for these languages (De Weerdt et al., 2003; Vermeerbergen, 2006; Baker et al., 2016). The fact that sign languages were —and often are —not treated equally to spoken languages impacts many areas in the lives of deaf people, for example language acquisition and education. Since 90 to 95 percent of deaf children are born to hearing parents and the use of sign language in the upbringing and education of deaf children was often not considered desirable (sometimes even detrimental), many deaf children acquired sign language from (slightly older) peers in school contexts, instead of their parents or teachers (Plann, 1997; Beelaert et al., 2009; Baker et al., 2016). In many cases, deaf children do not receive adequate early access to sign languages, and as a result experience delayed language access (Hall, 2017; Lillo-Martin and Henner, 2021). Thus, the importance of early intervention, visual communication and accessible language input with regards to deaf children is critical (Wille, 2021). In addition to using sign language, being surrounded by a majority spoken language also means most deaf people grow up to be functionally bilingual in both (at the very least to some degree) (Baker et al., 2016).

Comparing the lexicon of a spoken and sign language, it seems that the number of established or frozen lexemes of most sign languages, i.e. signs that are conventional in form and meaning, are smaller than with most spoken languages (Vermeerbergen, 2006). This may be due to several reasons. For example, when certain topics are not usually discussed in a language or in its community, e.g. when (certain) subjects are never taught in said language, it is possible that lexemes for certain specific concepts never naturally develop (Vermeerbergen, 1997). This may contribute to specific lexical gaps that can be observed for most sign languages. Another reason may be the lack of a standardised written form and the fact that video technology for sharing signs is relatively modern. However, it is important to note that signers have a wide variety of strategies at their disposal to fill these lexical gaps. Sign languages offer more flexibility when it comes to adapting and reusing signs in different ways and combining building blocks in novel ways to express nuance or new concepts, i.e. the productive lexicon as mentioned above.

No standardised writing system All these factors also play a role in how sign languages are documented. As there are no commonly accepted written forms of sign languages outside academic contexts, sign language resource creators encounter the challenge of how to represent information. Existing notation systems, such as HamNoSys (Hanke, 2004), used in research mainly focus on the exact phonetic representation of signs, but are too complex for both (quick) production and reading of longer passages.

As a compromise, researchers work with several keywords for dictionary translations or transcribe longer stretches of sign language data using *ID-glosses*, which represent each sign through a spoken language word that approximates its meaning, followed by a number or letter to distinguish different signs that share a meaning, e.g. the glosses TREE1, TREE2, TREE3 are used to transcribe different variations of signs with the meaning ‘tree’. The same sign is always transcribed using the same gloss. Unlike full translations, glosses lack some kinds of grammatical information, are context-independent and may deviate from the given meaning in context. Therefore glosses should never be seen as valid translations, as some information would be lost in translation, such as the actual number of signs and order in which they occur, even when the translation would require more or fewer words or a different word order. For example, a specific utterance in German Sign Language may be translated as “The cat is on the table” and be glossed as TABLE1 CAT3 INDEX1.

Additionally, most dictionary creation projects for sign languages have been solely focusing on established lexicons, missing out on other aspects, e.g. not taking productive signs into account. This does, however, not showcase the full richness or potential of a sign lexicon.

The lack of comprehensive documentation, paired with the scarcity of other high-quality sign language data, complicates further research and technical advancement like machine translation (MT) for sign languages.

2.3. Adoption and acceptance of technologies by deaf communities

With digital communication on the rise and new technological tools available, deaf individuals in Europe now have more opportunities to use language technologies for accessibility purposes and to translate information between spoken/written languages in a multilingual context. Deaf communities quickly adopt communication breakthroughs that offer accessibility and independence, such as teletypewriters, fax machines, text messaging, and videocalling, as these technologies allow direct communication without intermediaries.

Like many, deaf people also use text-to-text MT tools such as Google Translate[6] or DeepL[7] to translate between different languages, for either personal or professional purposes. Deaf persons also make creative use of some technologies, or in some cases, using them in ways other than the intended use, to match accessibility needs, for example:

- Transcription and automatic subtitling features, using those apps as a support for both accessibility and note-taking by saving a meeting transcription afterwards, to lighten the workload.
- Using built-in functions in other ways than were intended such as speech-to-text by asking an interlocutor to use the Dictation function in an iPhone in a note-taking app, even though this specific function was not intended to be used as a tool for accessibility but for the ease of hearing people.
- Other apps: Specific apps to be able to write text on a mobile device, either in a big font app or in a default note app, to make it easier to read for interlocutors, as an evolved method of writing back and forth on paper.

Adoption of technologies by deaf persons is influenced by a variety of factors, such as accessibility, affordability, availability and ease-of-use. These requirements also apply to hearing people. This demonstrates that contrary to popular belief, deaf individuals are not resistant to technologies. In fact, it is more accurate to note that most technologies have systematically excluded deaf people: for decades, many technological advancements have

[6] https://translate.google.com/
[7] https://www.deepl.com/translator

been increasingly centred around audio, leading to more barriers to participation for the deaf community. From telephones and radios to speakers and the latest technologies like Alexa, these devices are not accessible to those who are deaf. As a result, there is a growing need for technologies that cater to the unique communication needs of the deaf community, particularly in the realm of sign language recognition and translation.

It is worth noting that while numerous researchers are not deaf and do not have an in-depth understanding of sign language or deaf culture, they may still research technologies aimed at aiding deaf individuals. However, it is essential to acknowledge that a lack of familiarity with the specific linguistic and cultural aspects of sign languages and deaf communities can fail to consider these elements appropriately. Consequently, certain technologies may be perceived as ineffective or flawed by the deaf community, despite being praised by individuals who are not familiar with the challenges faced by deaf individuals.

There is a perfect illustration of this issue: Sign language gloves, claiming to translate sign language to text or speech in real-time, are not helpful for the deaf community (Erard, 2017). They fail to capture the full range of signs used in sign languages, including linguistic parameters beyond the hands, such as facial expressions. Moreover, they place the burden of communication on the deaf person who has to wear the glove to "fit into society". A similar project about a robotic arm developed by hearing students to"help" deaf people by fingerspelling words is not a sign language translation and only facilitates one-directional communication. Furthermore, they did not consider user-friendliness, as people do not want to haul around communication tools which are substantial in size.

To mitigate development issues such as these, it is important to prioritise a genuine and inclusive involvement of the deaf communities in the research process, as opposed to tokenistic involvement. This can help to avoid making assumptions about sign languages and deaf cultures, and can ensure that user-friendliness is a priority and thus encourage adoption of the technology. It is important for researchers to reflect on the reasoning and methods behind their projects. Some red flags for research and development, which are often named by deaf communities, to consider include:

- Adopting a paternalistic approach, such as aiming to "help fix" communication issues for deaf individuals.
- Holding biased views on how technology should be used by the deaf community, assuming that they simply need to adjust to the technology.
- Maintaining outdated or discriminatory beliefs about deaf communities and sign languages, including the use of discriminatory terminology such as "deaf and dumb" or referring to hearing people as "normal people".
- Design and development without consulting members of deaf communities/sign language users for a technology that is supposed to be for them or involving them only at the end of the project, leading to being unaware of what is really needed and following mistaken assumptions about what will suffice for accessibility.
- Neglecting to involve deaf experts, who play a crucial role in identifying the barriers that deaf individuals face with technologies and user-friendliness.
- Failing to involve sign language linguists who possess a deep understanding of the structure, grammar, and syntax of sign languages.

By avoiding these potential issues, researchers can better ensure that their projects are respectful, inclusive, and effective for the deaf communities. This way, deaf communities are open to embrace new (language) technologies that are designed for, and with, them. The right digital tools enable deaf persons to enjoy an increased accessibility and communication with a greater independence. At the same time, society increases its awareness and understanding of the unique needs and perspectives from deaf people, which will in turn lead to a greater inclusivity and a greater language equality for sign languages.

3. What is Language Technology?

Natural language[8] is the most common and versatile way for humans to convey information. We use language, our natural means of communication, to encode, store, transmit, share and process information. Processing language is a non-trivial, intrinsically complex task, as language is subject to multiple interpretations (ambiguity), and its decoding requires knowledge about the context and the world, while in tandem language can elegantly use different representations to denote the same meaning (variation).

The computational processing of human languages has been established as a specialised field known as *Computational Linguistics* (CL), *Natural Language Processing* (NLP) or, more generally, Language Technology (LT). While there are differences in focus and orientation, since CL is more informed by linguistics and NLP by computer science, LT is a more neutral term. In fact, LT is largely multidisciplinary in nature; it combines linguistics, computer science (and notably AI), mathematics and psychology among others. In practice, these communities work closely together, combining methods and approaches inspired by both, together making up *language-centric Artificial Intelligence*.

Language Technology is the multidisciplinary scientific and technological field that is concerned with studying and developing systems capable of processing, analysing, producing and understanding human languages, whether they are written, spoken or signed.

With its starting point in the 1950s with Turing´s renowned intelligent machine (Turing, 1950) and Chomsky´s generative grammar (Chomsky, 1957), LT enjoyed its first boost in the 1990s. This period was signalled by intense efforts to create wide-coverage linguistic resources, such as annotated corpora, thesauri, etc. which were manually labelled for various linguistic phenomena and used to elicit machine readable rules which dictated how language can be automatically analysed and/or produced. Gradually, with the evolution and advances in machine learning, rule-based systems have been displaced by data-based ones, i. e., systems that learn implicitly from examples. In the recent decade of 2010s we observed a radical technological change in NLP: the use of multilayer neural networks able to solve various sequential labelling problems. The success of this approach lies in the ability of neural networks to learn continuous vector representations of the words (or word embeddings) using vast amounts of unlabelled data and using only some labelled data for fine-tuning.

In recent years, the LT community has been witnessing the emergence of powerful new deep learning techniques and tools that are revolutionising the way in which LT tasks are approached. We are gradually moving from a methodology in which a pipeline of multiple modules was the typical way to implement LT solutions, to architectures based on complex neural networks trained with vast amounts of data, be it text, audio or multimodal. The success in these areas of AI has been possible because of the conjunction of four different research trends: 1) mature deep neural network technology, 2) large amounts of data (and for NLP processing large and diverse multilingual data), 3) increase in high performance computing (HPC) power in the form of GPUs, and 4) application of simple but effective self-learning approaches.

[8] This section has been provided by the editors. It is an adapted summary of Agerri et al. (2021) and of Sections 1 and 2 of Aldabe et al. (2021). It has been further adapted towards sign language technology by the authors of this paper.

For spoken languages, LT is trying to provide solutions for the following main application areas: Text analysis, speech processing, machine translation, information extraction and information retrieval, natural language generation, and human-computer interaction.

For sign languages specifically, the following areas are most relevant:

- **Sign Language Recognition** (information extraction from sign language) aims at enabling a computer to identify signs produced by sign language users allowing humans to communicate with electronic devices through sign language.
- **Sign Language Synthesis** aims at generating a signed message through a virtual signer or avatar.
- **Sign Language Translation** aims at the automatic translation from and to a sign language (in all possible combinations): from sign language to spoken/written language, from spoken/written language to sign language, as well as from sign language to another sign language.

LT for spoken languages is already fused in our everyday lives. As individual users we may be using it without even realising it, when we check our texts for spelling errors, when we use internet search engines or when we call our bank to perform a transaction. It is an important, but often invisible, ingredient of applications that cut across various sectors and domains. To name just very few, in the *health* domain, LT contributes for instance to the automatic recognition and classification of medical terms or to the diagnosis of speech and cognitive disorders. It is more and more integrated in *educational* settings and applications, for instance for educational content mining, for the automatic assessment of free text answers, for providing feedback to learners and teachers, for the evaluation of pronunciation in a foreign language and much more. In the *law/legal* domain, LT proves an indispensable component for several tasks, from search, classification and codification of huge legal databases to legal question answering and prediction of court decisions.

When aiming at equal facilities for all European languages by 2030 it is therefore of utmost importance that such technologies are being developed not only for spoken languages but also for Europe's sign languages.

The wide scope of LT applications evidences not only that LT is one of the most relevant technologies for society, but also one of the most important AI areas with a fast growing economic impact.[9]

4. Language Technology for Europe's Sign Languages

4.1. Language Data

The two main types of sign language resources are *lexical resources* and *corpora*. Lexical resources, such as dictionaries, describe individual signs, while corpora collect recordings of actual language use, such as signed conversations. Often, corpora supplement the recordings with annotations of what signs are being used and with translations into other languages.

[9] In a recent report from 2021, the global LT market was already valued at USD 9.2 billion in 2019 and is anticipated to grow at an annual rate of 18.4% from 2020 to 2028 (https://www.globenewswire.com/news-release/2021/03/22/2196622/0/en/Global-Natural-Language-Processing-Market-to-Grow-at-a-CAGR-of-18-4-from-2020-to-2028.html). A different report from 2021 estimates that amid the COVID-19 crisis, the global market for NLP was at USD 13 billion in the year 2020 and is projected to reach USD 25.7 billion by 2027, growing at an annual rate of 10.3% (https://www.researchandmarkets.com/reports/3502818/natural-language-processing-nlp-global-market).

For a detailed overview of relevant lexical and corpus resources, see the EASIER Report on Datasets for the Sign Languages of Europe (Kopf et al., 2021),[10] as well as the CLARIN resource family of sign language resources.[11]

4.1.1. Lexical Resources

Online dictionaries have been created for a number of sign languages. These lexical resources usually consist of a website containing short single sign videos linked to a specific gloss, or to a (written) translation of the sign in a spoken language. Due to the strong regional and dialectal variation in the vocabulary of many sign languages, one gloss/translation is often linked to multiple synonymous signs. Some dictionaries also provide definitions (in spoken or, less common, in signed language), examples or other types of information, although this is still rare. Due to the lack of established written forms for sign languages, most dictionaries can only be searched via the spoken language text of translations/glosses. Search tools aimed at features inherent to sign languages, such as handshapes and movement, are rare, and webcam-based video search is still in the stage of experimental prototypes.

During the process of corpus annotation, dictionaries can be helpful to identify signs and ensure that they are consistently annotated with the same ID-gloss. A common workflow is to link a lexical database containing the dictionary content directly to the annotations in the transcript. An example of this workflow is how the annotation software ELAN (Wittenburg et al., 2006)[12] can be linked to a Signbank[13] lexical database as its underlying vocabulary.

Example Dictionary In the following, the *Swedish Sign Language Dictionary (SSLD)* (Mesch et al., 2012; Svenskt teckenspråkslexikon, 2023) will be used as a case study to discuss various aspects of lexical sign language resources. The SSLD was created in 2008 at the University of Stockholm and includes over 20,000 public entries; it is a constantly growing resource. Signs can be searched via keywords in Swedish or English.

The SSLD is a rich resource with each entry providing detailed information for a sign (see Fig. 1). Every entry contains some or all of the following fields:[14]

- A video showing the citation form of the sign — as movement is an integral part of sign languages, static pictures are not an adequate way of depicting signs.
- A number of videos in Swedish Sign Language, providing a) usage examples, b) definitions and c) discussions of the origins of a sign.
- A phonetic definition of the sign, given both as a Swedish text description and as phonetic transcription using a notation specifically designed for Swedish Sign Language.
- The topics that are associated with the sign.
- Mouthings accompanying the sign, depicted as pictures and using abstract text labels.
- A unique ID to reference the entry unambiguously in other resources.
- The sign's gloss name used in the Swedish Sign Language Corpus *sts-korpus* (see Section 4.1.2).
- Number of occurrences of the sign in dictionary, corpus and survey materials, plus a link to a concordance view of sign occurrences in the *sts-korpus*.

[10] The contents of this report have also been integrated into the regularly updated Sign Language Dataset Compendium (Kopf et al., 2022a), located at https://www.sign-lang.uni-hamburg.de/lr/compendium/
[11] https://www.clarin.eu/resource-families/sign-language-resources
[12] https://archive.mpi.nl/tla/elan
[13] https://github.com/signbank/
[14] This information is documented at https://teckenspraklexikon.su.se/information/manual/sidan-enskilt-tecken

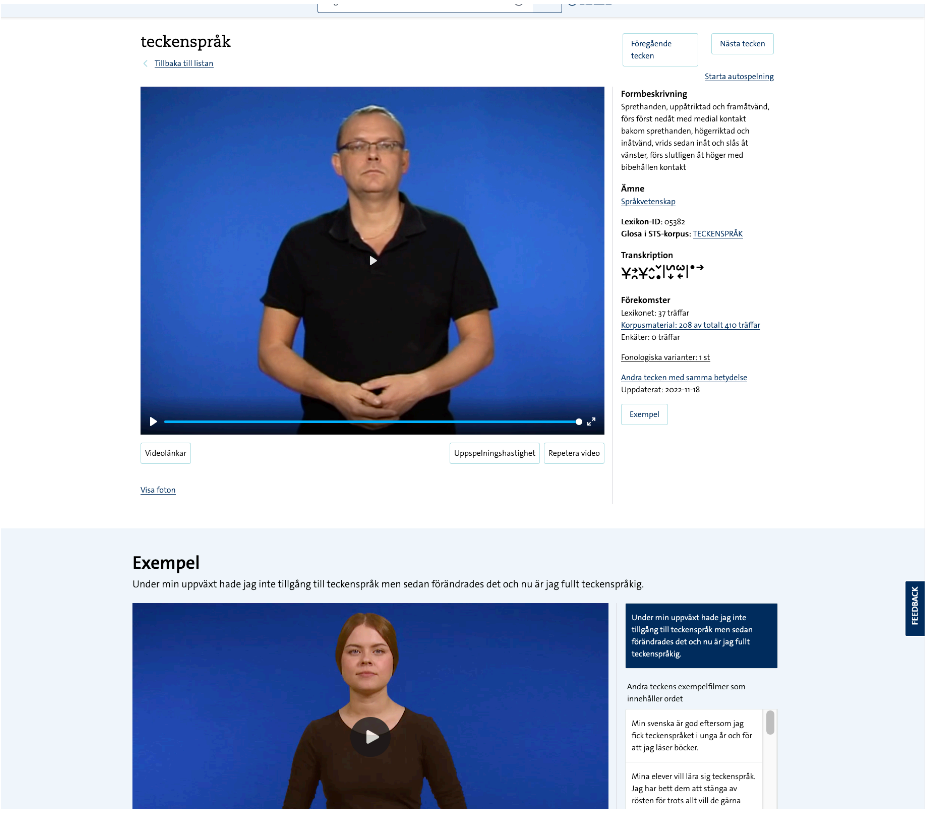

Figure 1: Entry in the Swedish Sign Language Dictionary for the sign meaning *sign language* (Swedish: *teckenspråk*). The top part shows a video of the base form of the sign plus a description, formal identifiers, phonetic transcription and corpus information. The bottom shows a usage example of the sign in a full sentence. Additional examples can be selected on the right.

- A link to phonological variants and signs with the same meaning.
- Information on when the entry was last updated.

As mentioned before, the SSLD is an unusually detailed resource that contains considerably more information than the majority of online dictionaries for sign languages.

Another form of lexical database are wordnets (Fellbaum, 1998), which are semantic networks providing lexico-semantic relations between the words or signs of a language. Information is structured based on concepts, rather than words, assigning each word/sign to all concepts it can represent, grouping it together with other words/signs that represent the same concept. In some cases, cross-compatibility for wordnets of different languages is established through interlingual indices that establish equivalence between concept entries (Bond and Foster, 2013; Bond et al., 2016). For sign languages, wordnets have been considered for a while (Ebling et al., 2012; Shoaib et al., 2014), but little data has yet been published. At the time of writing, small datasets for Greek, British and German Sign Language have been released (Bigeard et al., 2022), while datasets on Flemish Sign Language (Schuurman et al., 2023) and American Sign Language (Lualdi et al., 2021) are still in production.

Another form of lexical resource, one which is important in sign language synthesis, is a lexicon of glosses and the description of the associated signs in a formal language that allows to drive the movements of an avatar, such as SigML (Kaur and Kumar, 2016). Only few of such resources are available. An example for Sign Language of the Netherlands is Esselink et al. (2022).

4.1.2. Corpora

The term 'corpus' refers to a slightly different concept in sign language linguistics than in spoken language linguistics. Modern linguistic corpora are machine-readable and maximally representative of the language and its users. Sign language linguists refer to any collection of video recordings that they base their analysis on a *corpus*, most of them too small to be representative and/or machine-readable (Fenlon et al., 2015). This is partly due to the lack of a written form for signed languages (see Section 2.2) which means that video recordings are the only way to accurately represent signed content. These recordings are usually accompanied by gloss transcriptions and/or translations into spoken languages, although the substantial labour cost of creating such annotations means that most corpora can only provide them for parts of their data.

The level of detail and amount of linguistic factors considered within annotations also vary widely across corpora. As shown in Kopf et al. (2022b), manual signs are covered by almost all annotations, non-manuals are annotated less frequently and grammatical aspects, such as part of speech (POS) or dependency syntax, can seldom if ever be found, in part because for some phenomena like POS, no consensus has yet been reached for their theoretical definition. As there are few commonly agreed annotation standards, transcripts of various corpora differ in their approach (De Sisto et al., 2022). One commonly established best practice is to use ID-glosses (see Section 2.2) and link them to a lexical database.

Sign language data are often accompanied by translations into the spoken language of the same geographical area. These are full *translations* between two natural languages, and not a *transcript* of the sign language: There is no one-to-one alignment between individual signs and their translations, and the two languages have different grammars, so alignment between translation equivalents is only possible at the utterance level.

As sign language corpora depend on the use of video recordings that show their participants, all corpus data are sensitive and must take participant privacy into account. Full anonymisation while retaining sufficient language information is not possible, although anonymisation of personally identifiable information can be achieved by, e.g. blacking out critical parts (Isard, 2020). Dataset creators must take into account their ethical and legal responsibilities towards their participants (Harris et al., 2009; Crasborn, 2010; De Meulder, 2021), which by definition are part of a linguistic (and usually cultural) minority group.

Most linguistic corpora offer open access for at least parts of the available data; for some of them registration or an individual license agreement is required (issues of acquiring sign language datasets are further discussed in De Sisto et al. (2022)).

Example Corpus As an example of a sign language corpus, consider the *Swedish Sign Language Corpus.*[15] It was collected as part of the STS Corpus project, based at Stockholm University. The corpus consists of 25 hours of semi-spontaneous dialogues and narratives by 42 informants. The informants are from three regions in Sweden and aged between 20 and 82. By the end of the three-year project, 14% of the corpus material —corresponding to 2 hours and 30 minutes —had been annotated with glosses and a translation into Swedish. The annotation comprised 3,600 different signs and approximately 25,500 tokens (Mesch et al., 2012).

The corpus can be viewed through the Swedish-language online interface *STSkorpus*,[16] which provides a transcript viewer and a text-based search tool. Users can search for glosses

[15] https://www.su.se/english/research/research-projects/swedish-sign-language-corpus
[16] https://teckensprakskorpus.su.se

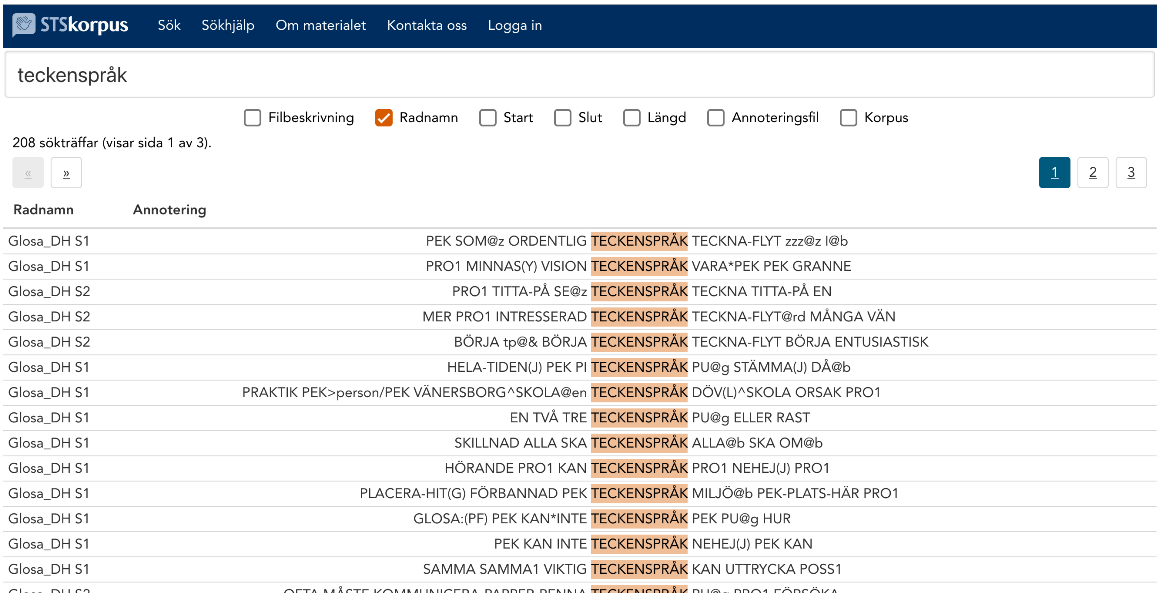

Figure 2: STSkorpus search result for "teckenspråk" (sign language). Matched glosses are shown as a concordance view with neighbouring glosses.

in the corpus transcripts. Search results are shown as a concordance view, i.e. each match is shown in the context of its neighbouring glosses, to help users select a relevant transcript (see Fig. 2). Selecting a match opens its transcript at the corresponding point in the recording. The transcript viewer shows the recordings as well as annotation tiers for glosses and translations. Separate tiers are provided for each signer as well as for which hand they produced a sign with (see Fig. 3). As the recording is played, the annotation scrolls along to match the video.

STSkorpus is also connected to Swedish Sign Language Dictionary (see Section 4.1.1). Clicking on a gloss in the transcript viewer pops up basic dictionary information for that sign (video, description, phonetic transcription) and allows users to move to its full entry in the dictionary or search STSkorpus for other occurrences.

As can be seen through the example of the Swedish Sign Language Corpus, sign language linguistic corpora are very rich in the data they provide, but very small compared to many spoken language corpora. Additionally, they are mostly multilingual in the sense that target and meta language are not the same, but usually cover only one sign language. Notable exceptions are Dicta-Sign (Efthimiou et al., 2012) and ECHO (Brugman et al., 2004) that cover multiple sign languages.

4.1.3. Quality of the sign language data

Most of the sign language corpora, dubbed as e.g. the Corpus VGT, the BSL Corpus, the Corpus NGT, and the DGS Corpus mainly consist of data produced by members of the deaf community for whom their sign language is their primary or preferred language. In the context of spoken

Figure 3: STSkorpus transcript viewer, showing a recording and its annotations. When played, the annotations move along to match the video.

languages, we would consider these speakers as *native*[17] speakers or *L1* speakers, but for sign language we use *authentic*.

Another set of sign language resources is available that consists of data in which sign language is the target language, i.e. the result of a translation or interpretation from a spoken language source. This causes the sign language produced to be influenced by the source language, hence to be different from spontaneously articulated utterances, also known as *translationese* (Graham et al., 2020). In addition, often sign language interpretation is performed by hearing interpreters for whom the sign language is a second language; this can affect the quality of the data, since the language produced by a second language signer, just as it is for a second language speaker, cannot be comparable to that by an authentic signer. Moreover, interpretation often takes place simultaneously and under time constraints implying that interpreters often need to choose for conveying the message in the quickest and most efficient way over providing the most correct translation of the source speech. This is commonly the case for e.g. TV broadcasts and parliament plenary sessions that have been interpreted into a sign language. While such data may be available in larger quantities, its usefulness and effect on the quality of LT is still under debate.

4.2. Language Technologies and Tools

Unlike spoken language annotation, which these days can often be supported by advanced language technologies like automated speech recognition or machine translation, technologies for sign languages are still considerably more limited. While progress has been made

[17] In the case of sign language, using the term *native* can be quite problematic (see Costello et al. (2008); *authentic* refers to those cases in which a sign language was learnt at a later stage in life, e.g. school age, but does constitute the main and preferred manner of communication for the person.

in recent years, e.g. towards the creation of (semi-)automatic sign segmentation (Mukushev et al., 2022; Woll et al., 2022), most sign language technologies do not yet meet the quality requirements of corpus creators.

In turn, to approach those quality requirements, most sign language technologies require considerable amounts of annotated training data that are not yet available for *any* sign language. Many systems are therefore limited to proof-of-concept demonstrations in specialised domains with particularly limited linguistic variation, such as weather reports.

In the context of machine translation between spoken languages, typically one would require millions of parallel sentences to achieve decent quality. One of the most commonly used corpora for machine translation, Europarl (Koehn, 2005), contains approximately 2 million parallel sentences for the high-resource language pairs and around 500 thousands for the low-resource ones. On the extreme side, consider the work of Hassan et al. (2018), which presents a model trained on more than 25 million parallel sentences. However, the DGS Corpus (Prillwitz et al., 2008) — one of the largest annotated sign language corpora — encompasses 50 hours of publicly available data (video material and full transcriptions) which correspond to around 60,000 parallel sentences.

The Sign-Lang @ LREC Anthology keeps track of LT tools that have been described in publications at the Language Resource and Evaluation conferences.[18]

Lessons learned from past research projects inform us that the successful implementation of a project within this domain, must be bolstered by close engagement with deaf communities, and co-construction of MT agendas. Including deaf experts on MT project teams is also essential (Murtagh et al., 2021).

4.2.1. Sign Language Recognition (SLR)

This is the task of recognising and understanding the meaning of signs. Each sign needs to be assigned a label, which is typically done using glosses (Núñez-Marcos et al., 2023). A distinction can be made between isolated SLR and continuous SLR. In the former, one video is assigned exactly one label and the task is to predict which sign is present in the video. In the latter, every video contains multiple signs (as part of continuous signing) and the task is to predict the labels for all the signs in the video in their order of appearance.

One application of isolated SLR is sign language dictionary lookup (Hassan et al., 2021). Continuous SLR can be applied as part of an SLT pipeline (Camgöz et al., 2020). In the latter, the SLR model first extracts information (in terms of glosses or embeddings) from the sign language videos, and the MT model then translates this information into text in another language.

Rastgoo et al. (2021) provide a survey of SLR. As a video-based task, SLR is typically tackled with deep learning. In both isolated and continuous SLR, the video is first processed using a vision model (typically a convolutional neural network or vision transformer) to extract features. These features can be latent representations of the network (with limited interpretability), or they can be more structured in the form of keypoints. Keypoints are extracted using human pose estimation tools such as OpenPose (Cao et al., 2021) or BlazePose (Bazarevsky et al., 2020), and these tools predict 2D or 3D Cartesian coordinates for every joint and other important landmark, e.g., the eyes and ears, in the human body. Despite the increased interpretability of keypoints compared to arbitrary latent representations, they are less popular in SLR because they lack robustness when applied to sign language data (Moryossef et al., 2021).

The state of the art in SLR is continuously evolving. Most of the research into isolated SLR is performed on the WLASL (Li et al., 2020), AUTSL (Sincan and Keles, 2020) and MS-ASL (Joze and Koller, 2019) datasets, which contain recordings of individual signs. A more challenging

[18] https://www.sign-lang.uni-hamburg.de/lrec/tool/

and realistic approach to isolated SLR is to use signs cut from continuous signing, which features coarticulation and sign transitions. Such an approach is more challenging (De Coster et al., 2020) and requires more elaborate techniques (Albanie et al., 2020). For continuous SLR, the de facto benchmark is the RWTH-PHOENIX-Weather-2014T (Camgöz et al., 2018) dataset. Similar techniques are applied as for isolated SLR, but a different optimisation algorithm is used because the goal is to predict multiple signs in order. This algorithm is called connectionist temporal classification (Graves, 2012) and it is also used in automatic speech recognition.

4.2.2. Sign Language Synthesis

Due to the visual-gestural modality difference and the fact that sign language has no standardised and commonly used written form, technology leverages the use of avatars in order to communicate a sign language utterance.

Recent research with regard to user acceptance outlines the importance of the fluency of movement of the avatar and the quality of the avatar itself, with regard to the successful synthesis of sign language (Quandt et al., 2022). Further factors, aside from the uncanny valley issue (Diel and MacDorman, 2021), that aggravate the use of avatar technology, include the size of hands, the unnatural movement of the hands, shoulders or head and the lack of inclusion of various facial features, such as the eyes, eyebrows, lips etc. (Kipp et al., 2011). These are essential in the communication of various linguistic information, for example indicating negation, if something is a question, or topic-phrases (Murtagh, 2019). Animation issues can also occur when the hands must overlap varying parts of the avatar body or one hand overlaps another hand.

Another challenge in relation to sign language Synthesis is the sign language lexicon-animation interface. Scripting languages have been developed to bridge this gap including Web3D open standards such as Virtual Reality Mark Up Language (VRML) and X3D (Su and Furuta, 1998; Grieve-Smith, 2002; Papadogiorgaki et al., 2005; Yu and Lu, 2013). SiGML (Kipp et al., 2011; Neves et al., 2020), and BML (Murtagh et al., 2022) are among other scripting languages, which were developed to bridge this interface. However, there is no current standard markup language that has been agreed upon.

The development of avatars to generate sign language has recently expanded to commercial products, which are albeit offered in a limited spectrum of use cases, and they seem to be early stage products or even in a development phase. Since they are not publicly available, little is known to which extent they are functional. One example is *SIMAX*,[19] provided by the Austrian company SignTime. It is software for translating text in 3D-animated sign language. It operates with an actively learning databank on the background, whose output is post-produced by sign language experts. *Charamel*,[20] a company providing avatar assistants, has been constructing signing avatars for the train announcements of the German railway (as part of the AVASAG project, funded by BMBF), whereas another avatar service has been constructed for a communal participation project. *JASigning* (Van Gemert et al., 2022) is a similar research project, aiming to provide a signing avatar for travel announcements of the railway in the Netherlands. The British company *Synapse*[21] is accomplishing a similar project for the railway station of Huddersfield, UK. Our research also led to some avatar products that despite having being announced, had a very short life and are not available any more (e.g. Sign-360 by the French company 4mocaplab),[22] possibly indicating the difficulty that this task poses for the commercial sector.

[19] https://simax.media
[20] https://gebaerdensprach-avatar.charamel.de
[21] https://www.signapse.ai
[22] https://www.4mocaplab.com/fr/projects/sign-360/

4.2.3. Sign Language Translation (SLT)

Núñez-Marcos et al. (2023) and De Coster et al. (2022) present recent surveys on SLT. The latter focuses only on translating sign language from video into a spoken language text.

Many systems use a pipeline technique, i.e. first applying SLR, and then translating the recognised signs into spoken language text. This text can then be translated into another spoken language, if required, using text-to-text MT systems. The spoken language text can then be turned into a series of signs using sign language synthesis. The output of the SLR in this pipeline does not need to be a label or a gloss, but it can be a multidimensional continuous representation, similar to the encoder output in textual MT systems. However, if the input of the encoder consists of raw video without linguistic information, then the representations that are generated in the encoder do not capture the syntax or the semantics of the sign language, and the translation model is forced to learn both the sign language semantics and the translation at the same time. If linguistic / semantic information is used when encoding the video, then only the translation needs to be learned. Linguistic information can be added by augmenting the input video using sign language recognition features, such as glosses or another written representation of signs or features of these signs, such as specific movements. State of the art systems seem to currently mainly focus on using glosses (De Coster et al., 2022). Jointly training SLR and SLT systems to generate both glosses and translations from videos can help to include such linguistic knowledge in the encoder (Camgöz et al., 2020).

An alternative to the pipeline approach are end-to-end systems that do not make use of gloss annotations or any other linguistic information about the sign language. The creation of such systems is currently hampered by the limited availablity of data. Therefore, such systems are often outperformed by the above mentioned pipeline systems (De Coster et al., 2022).

When starting from a spoken language and translating into an sign language, a similar pipeline approach can be followed (Stoll et al., 2020): in a first step encode the spoken language into a series of glosses and then lookup the glosses in a dictionary of motion, which is a sequence of skeletal poses, built using keypoint extracting techniques, as discussed above. Then, the video is generated based on these skeletal motions.

Saunders et al. (2020) present an end-to-end approach from spoken language text to a sequence of 3D sign poses that could directly be used to animate an avatar.

A demonstration system translating from several spoken languages into skeleton motion in several sign languages is available at https://research.sign.mt (Moryossef and Goldberg, 2021), which also provides an elaborate description of sign language processing.

4.3. Projects, Initiatives, Stakeholders

The European sign language communities are represented by EUD,[23] the European Union of the Deaf, which represents 31 national associations of the deaf, from all of the 27 EU countries, plus Iceland, Norway, Switzerland and the United Kingdom. Appendix A presents a list of the EUD members with their sign languages, and estimated number of signers. For more information on sign language legislation we refer to Wheatley and Pabsch (2012).

Sign language interpreters in Europe are represented by EFSLI,[24] the European Forum of Sign Language Interpreters, which consists of national and regional associations with individual and associate members. EFSLI is working towards the higher status of the profession of sign language interpreters in Europe.

[23] https://www.eud.eu
[24] https://efsli.org

The Sign Language Linguistics Society[25] promotes sign language research on an international scale, and there are also national research centres that focus on specific sign languages. A map of places where sign languages are studied is available at https://www.google.com/maps/d/edit?mid=1spPrlstOOHIkUtQpjRdH9RlSF1Y&usp=sharing.

SignON[26] and EASIER[27] are the two current projects funded by the EU call "An empowering, inclusive Next Generation Internet"[28] which was specifically addressed towards research into sign language translation technology. In previous EU calls we see funding of, amongst others the SignHub project,[29] Content4All,[30] and Dicta-Sign.[31] For an exhaustive list of European funded projects concerning sign language, we refer to https://cordis.europa.eu/search?q='sign'%20AND%20'language'.

5. Cross-Language Comparison

The papers in the ELE series about spoken languages contain a section in which these spoken languages are compared to each other with respect to the available tools and resources. For that section we refer to any of these ELE papers.[32] According to that section all sign languages fall into the category named *weak/no support*.

Instead of repeating that section, we have opted to discuss what is available for the different sign languages of Europe.

5.1. Levels of Technology Support

Figure 4 (Morgan et al., 2022) outlines, based on a report on sign language datasets (Kopf et al., 2021), that for the majority of European sign languages no high-quality training data in the form of corpora or lexical resources exist.[33] For approximately half of them small and fragmented datasets can be found. This means that for almost half of the European SLs no suitable resources that can be used with language technology are available (see Fig. 4).

5.2. European Language Grid as Ground Truth

At the time of writing (December 2022), the ELG catalogue comprises more than 11,500 metadata records, encompassing both data and tools/services, covering almost all European languages – both official and regional/minority ones. The ELG platform harvests several major LR/LT repositories[34] and, on top of that, more than 6,000 additional language resources and tools were identified and documented by language informants in the ELE consortium. These records contain multiple levels of metadata granularity as part of their descriptions.

Relating to sign languages, in the ELG, in December 2022 we found 61 sign language resources. Figure 5 shows how they are distributed over different sign languages. Of all sign

[25] https://slls.eu/
[26] https://signon-project.eu
[27] https://www.project-easier.eu
[28] https://cordis.europa.eu/programme/id/H2020_ICT-57-2020
[29] https://ww3.thesignhub.eu
[30] https://doi.org/10.3030/762021
[31] https://www.sign-lang.uni-hamburg.de/dicta-sign/portal/
[32] Available at https://european-language-equality.eu/deliverables/
[33] Note that certain size and quality criteria had to be met in order for datasets to be listed in this report.
[34] At the time of writing, ELG harvests ELRC-SHARE, LINDAT/CLARIAH-CZ, CLARIN.SI, CLARIN-PL and HuggingFace.

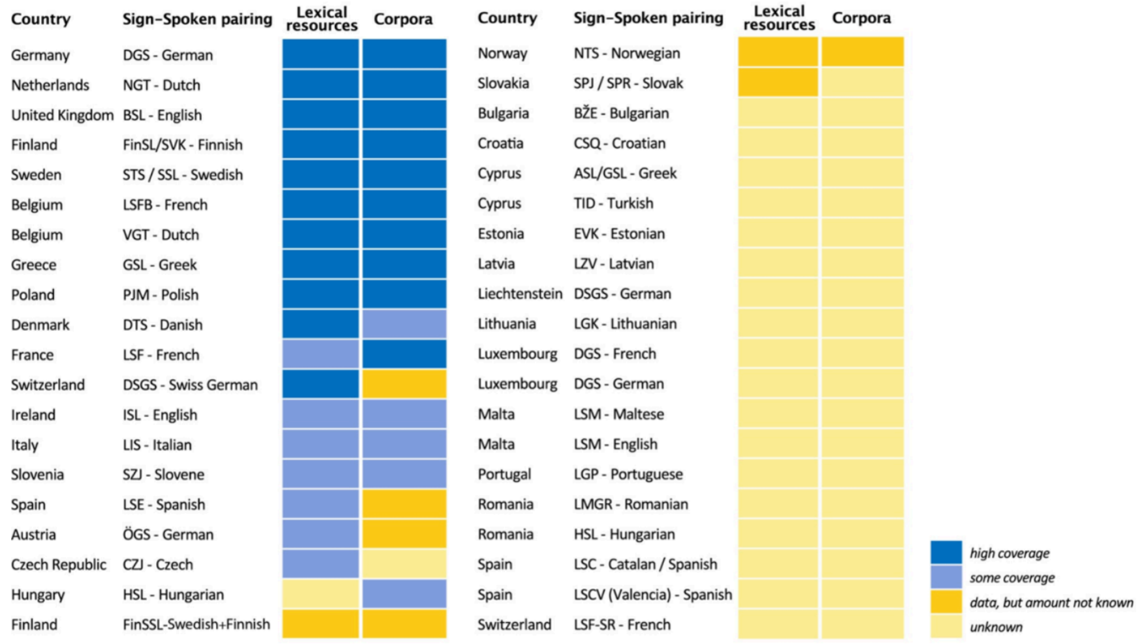

Figure 4: Chart by Morgan et al. (2022) on the state of European sign language resources.

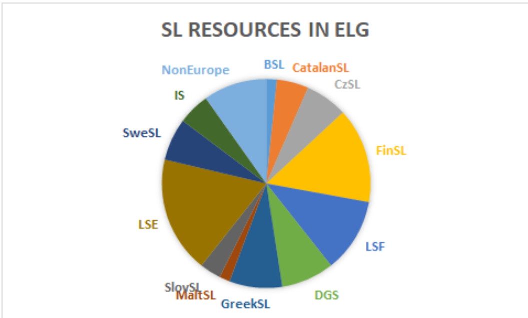

Figure 5: Distribution of sign language resources in the ELG.

language records, there were three on International Sign,[35] and six on non-European sign languages. The remaining 52 are distributed over 11 SLs. The three most represented SLs are Spanish, Finnish and French Sign Language, which together represent more than 50% of all European sign language resources in ELG.

It is clear that due to the low absolute numbers of these resources, it is hard to draw any conclusions with respect to how strong these SLs are compared to one another. It is also clear that ELG is missing metadata to several sign language datasets which are identified in, amongst other things, the sign language compendium.

[35] It needs to be mentioned that International Sign is not to be considered a natural sign language (Mesch, 2010), but rather a "mode of communication" which is used by signers who do not have a shared sign language (Hidding and Crasborn, 2011); even though it is based on a predefined code, it is highly context-dependent: it strongly uses iconicity and pantomimic structures, and exploits elements from the sign languages of the people communicating (Hidding and Crasborn, 2011).

6. Summary and Conclusions

Due to progress in artificial intelligence, development of language technologies for sign languages have become feasible. The deep learning paradigm has led to impressive progress in fields such as computer vision, natural language processing and machine translation. The gaming industry has booked immense progress on virtual agents.

There is one major downside to the deep learning paradigm, which is that it is more "data-hungry" then any other previous machine learning paradigm. So, in order to get language technology tools with a decent performance, huge amounts of data are required. It is clear that for sign language technologies, the required amounts of data are not available. Furthermore, the effect of the quality of the data which is available for training the AI systems on the output quality of the tools remains unclear.

If Europe wants to achieve language equality by 2030, and if this includes Europe's sign languages, it is of utmost importance that large sign language infrastructure projects are funded in which the deaf communities, deaf experts and sign language linguists cooperate with language technologists, computer vision researchers and virtual agent specialists co-operate on creating sufficiently large high quality multilingual sign language corpora which can be used with relative ease by the AI researchers. These projects should not only consist of building sign language datasets, but should also extensively study the effect of data size, quality and other independent variables on translation quality and user acceptance.

A renewed call for projects similar to the call that funded both the SignON and EASIER projects would therefore be a minimum, complemented with national infrastructure projects.

References

Rodrigo Agerri, Eneko Agirre, Itziar Aldabe, Nora Aranberri, Jose Maria Arriola, Aitziber Atutxa, Gorka Azkune, Arantza Casillas, Ainara Estarrona, Aritz Farwell, Iakes Goenaga, Josu Goikoetxea, Koldo Gojenola, Inma Hernaez, Mikel Iruskieta, Gorka Labaka, Oier Lopez de Lacalle, Eva Navas, Maite Oronoz, Arantxa Otegi, Alicia Pérez, Olatz Perez de Viñaspre, German Rigau, Jon Sanchez, Ibon Saratxaga, and Aitor Soroa. European Language Equality D1.2: Report on the state of the art in Language Technology and Language-centric AI, September 2021. URL https://european-language-equality.eu/wp-content/uploads/2021/10/ELE_Deliverable_D1_2.pdf.

Samuel Albanie, Gül Varol, Liliane Momeni, Triantafyllos Afouras, Joon Son Chung, Neil Fox, and Andrew Zisserman. BSL-1K: Scaling up co-articulated sign language recognition using mouthing cues. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*, pages 35–53. Springer, 2020. doi: 10.1007/978-3-030-58621-8_3.

Itziar Aldabe, Georg Rehm, German Rigau, and Andy Way. European Language Equality D3.1: Report on existing strategic documents and projects in LT/AI, November 2021. URL https://european-language-equality.eu/wp-content/uploads/2021/12/ELE___Deliverable_D3_1__revised_.pdf.

Anne Baker, Beppie van den Bogaerde, and Roland Pfau. *The Linguistics of Sign Languages: An Introduction*. Not in. John Benjamins Publishing Company, 2016. ISBN 9789027212313. URL https://books.google.be/books?id=9Cn4jgEACAAJ.

Valentin Bazarevsky, Ivan Grishchenko, Karthik Raveendran, Tyler Zhu, Fan Zhang, and Matthias Grundmann. Blazepose: On-device real-time body pose tracking. arXiv preprint, 2020. URL https://doi.org/10.48550/arXiv.2006.10204.

Bram Beelaert, Christine Bruyneel, and Kaat Leeman. *Vive la parole? Milaan 1880 als scharniermoment in het dovenonderwijs*. Fevlado-Diversus, 2009.

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

Francis Bond and Ryan Foster. Linking and Extending an Open Multilingual Wordnet. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1352–1362, Sofia, Bulgaria, August 2013. Association for Computational Linguistics. URL https://aclanthology.org/P13-1133.

Francis Bond, Piek Vossen, John McCrae, and Christiane Fellbaum. CILI: the Collaborative Interlingual Index. In *Proceedings of the 8th Global WordNet Conference (GWC)*, pages 50–57, Bucharest, Romania, 27–30 January 2016. Global Wordnet Association. URL https://aclanthology.org/2016.gwc-1.9.

Mary Brennan. *Word formation in BSL*. PhD thesis, Department of Linguistics, Stockholm University, 1990. URL https://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-63586.

Hennie Brugman, Onno Crasborn, and Albert Russel. Collaborative Annotation of Sign Language Data with Peer-to-Peer Technology. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal, May 2004. European Language Resources Association (ELRA). URL http://www.lrec-conf.org/proceedings/lrec2004/pdf/473.pdf.

Necati Cihan Camgöz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. Neural sign language translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7784–7793, 2018. doi: 10.1109/CVPR.2018.00812.

Necati Cihan Camgöz, Oscar Koller, Simon Hadfield, and Richard Bowden. Sign language transformers: Joint end-to-end sign language recognition and translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10023–10033, 2020. doi: 10.1109/CVPR42600.2020.01004.

Zhe Cao, Gines Hidalgo Martinez, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1):172–186, 2021. doi: 10.1109/TPAMI.2019.2929257.

Noam Chomsky. *Syntactic structures*. The Hague: Mouton, 1957.

Brendan Costello, Javier Fernández, Alazne Landa, and Ronice Müller de Quadros. The non- (existent) native signer: sign language research in a small deaf population. In *Sign Languages: spinning and unraveling the past, present and future. TISLR9, forty five papers and three posters from the 9th. Theoretical Issues in Sign Language Research Conference, Florianopolis, Brazil, 2006*, pages 77–94, Florianopolis, Brazil, 2008. Pétropolis/RJ, BrazilEditora Arara Azul. ISBN 9788589002270. URL https://hdl.handle.net/11245/1.317826.

Onno Crasborn. What Does "Informed Consent" Mean in the Internet Age? Publishing Sign Language Corpora as Open Content. *Sign Language Studies*, 10(2):276–290, 2010. ISSN 0302-1475. URL https://www.jstor.org/stable/26190572.

Mathieu De Coster, Mieke Van Herreweghe, and Joni Dambre. Sign language recognition with transformer networks. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 6018–6024, Marseille, France, May 2020. European Language Resources Association. ISBN 979-10-95546-34-4. URL https://aclanthology.org/2020.lrec-1.737.

Mathieu De Coster, Dimitar Shterionov, Mieke Van Herreweghe, and Joni Dambre. Machine translation from signed to spoken languages: State of the art and challenges. *CoRR*, abs/2202.03086, 2022. URL https://arxiv.org/abs/2202.03086.

Maartje De Meulder. Is "good enough" good enough? Ethical and responsible development of sign language technologies. In *Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL)*, pages 12–22. Association for Machine Translation in the Americas, 2021. URL https://aclanthology.org/2021.mtsummit-at4ssl.2.

Mirella De Sisto, Vincent Vandeghinste, Santiago Egea Gómez, Mathieu De Coster, Dimitar Shterionov, and Horacio Saggion. Challenges with Sign Language Datasets for Sign Language Recognition and Translation. In *Proceedings of the Language Resources and Evaluation Conference*, pages 2478–2487, Marseille, France, June 2022. European Language Resources Association. URL https://aclanthology.org/2022.lrec-1.264.

Kristof De Weerdt, Eline Vanhecke, Mieke van Herreweghe, and Myriam Vermeerbergen. *De dovengemeenschap in Vlaanderen: doorlichting, sensibilisering en standaardisering van de Vlaamse gebarentaal: Eindrapport, luik 2. Op onderzoek naar de Vlaamse gebaren-schat.* Fevlado-Diversus, 2003.

Alexander Diel and Karl F. MacDorman. Creepy cats and strange high houses: Support for configural processing in testing predictions of nine uncanny valley theories. *Journal of Vision*, 21(4), 2021. doi: 10.1167/jov.21.4.1.

Sarah Ebling, Katja Tissi, and Martin Volk. Semi-Automatic Annotation of Semantic Relations in a Swiss German Sign Language Lexicon. In Onno Crasborn, Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Jette Kristoffersen, and Johanna Mesch, editors, *Proceedings of the LREC2012 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon*, pages 31–36, Istanbul, Turkey, May 2012. European Language Resources Association (ELRA). URL https://www.sign-lang.uni-hamburg.de/lrec/pub/12010.pdf.

Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, John Glauert, Richard Bowden, Annelies Braffort, Christophe Collet, Petros Maragos, and François Lefebvre-Albaret. Sign language technologies and resources of the Dicta-Sign project. In Onno Crasborn, Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Jette Kristoffersen, and Johanna Mesch, editors, *Proceedings of the LREC2012 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon*, pages 37–44, Istanbul, Turkey, May 2012. European Language Resources Association (ELRA). URL https://www.sign-lang.uni-hamburg.de/lrec/pub/12025.pdf.

Michael Erard. Why Sign Language Gloves Don't Help Deaf People. *The Atlantic*, 11, 2017. URL https://www.theatlantic.com/technology/archive/2017/11/why-sign-language-gloves-dont-help-deaf-people/545441/.

Lyke Esselink, Floris Roelofsen, Jacub Dotlacil, Shani Mende-Gillings, Maartje De Meulder, Nienke Sijm, and Anika Smeijers. Exploring automatic text-to-sign translation in a healthcare setting. Submitted to Universal Access in the Information Society, 2022. URL https://www.signlab-amsterdam.nl/publications/Esselink22.pdf.

Christiane Fellbaum, editor. *WordNet: An Electronic Lexical Database*. The MIT Press, 1998. ISBN 978-0-262-27255-1. doi: 10.7551/mitpress/7287.001.0001.

Jordan Fenlon, Adam Schembri, Trevor Johnston, and Kearsy Cormier. *Documentary and Corpus Approaches to Sign Language Research*, chapter 10, pages 156–172. John Wiley & Sons, Ltd, 2015. ISBN 9781118346013. doi: 10.1002/9781118346013.ch10.

Yvette Graham, Barry Haddow, and Philipp Koehn. Statistical power and translationese in machine translation evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 72–81, Online, 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.6.

Alex Graves. Connectionist temporal classification. In Alex Graves, editor, *Supervised sequence labelling with recurrent neural networks*, volume 385 of *Studies in Computational Intelligence*, pages 61–93. Springer, Berlin, 2012. doi: 10.1007/978-3-642-24797-2_7.

Angus B. Grieve-Smith. SignSynth: A Sign Language Synthesis Application Using Web3D and Perl. In Ipke Wachsmuth and Timo Sowa, editors, *Gesture and Sign Language in Human-Computer Interaction*, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, 2002. doi: 10.1007/3-540-47873-6_14.

W. C. Hall. What you don't know can hurt you: The risk of language deprivation by impairing sign language development in deaf children. *Maternal and Child Health Journal*, 21(5):61—965, 2017. doi: 10.1007/s10995-017-2287-y.

Thomas Hanke. HamNoSys – representing sign language data in language resources and language processing contexts. In Oliver Streiter and Chiara Vettori, editors, *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*, pages 1–6, Paris, France, 2004. European Language Resources Association (ELRA). URL https://www.sign-lang.uni-hamburg.de/lrec/lrec/pubs/04001.pdf.

Raychelle Harris, Heidi M. Holmes, and Donna M. Mertens. Research ethics in sign language communities. *Sign Language Studies*, 9(2):104–131, 2009. ISSN 0302-1475. URL https://www.jstor.org/stable/26190667.

Hany Hassan, Anthony Aue, Chang Chen, Vishal Chowdhary, Jonathan Clark, Christian Federmann, Xuedong Huang, Marcin Junczys-Dowmunt, William Lewis, Mu Li, Shujie Liu, Tie-Yan Liu, Renqian Luo, Arul Menezes, Tao Qin, Frank Seide, Xu Tan, Fei Tian, Lijun Wu, Shuangzhi Wu, Yingce Xia, Dongdong Zhang, Zhirui Zhang, and Ming Zhou. Achieving Human Parity on Automatic Chinese to English News Translation. arXiv preprint, 2018. URL https://doi.org/10.48550/arXiv.1803.05567.

Saad Hassan, Oliver Alonzo, Abraham Glasser, and Matt Huenerfauth. Effect of sign-recognition performance on the usability of sign-language dictionary search. *ACM Transactions on Accessible Computing (TACCESS)*, 14(4):1–33, 2021. doi: org/10.1145/3470650.

Anja Hidding and Onno Crasborn. Signed languages and globalization. *Language in Society*, 40(4): 483–505, 2011. doi: 10.1017/S0047404511000480.

Amy Isard. Approaches to the anonymisation of sign language corpora. In Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Julie A. Hochgesang, Jette Kristoffersen, and Johanna Mesch, editors, *Proceedings of the LREC2020 9th workshop on the representation and processing of sign languages: Sign language resources in the service of the language community, technological challenges and application perspectives*, pages 95–100, Paris, France, 2020. European Language Resources Association (ELRA). ISBN 979-10-95546-54-2. URL https://www.sign-lang.uni-hamburg.de/lrec/lrec/pubs/20037.pdf.

Hamid Reza Vaezi Joze and Oscar Koller. MS-ASL: A Large-Scale Data Set and Benchmark for Understanding American Sign Language. In Kirill Sidorov and Yulia Hicks, editors, *Proceedings of the British Machine Vision Conference (BMVC)*, pages 41.1–41.16. BMVA Press, 2019. doi: 10.5244/C.33.41.

Khushdeep Kaur and Parteek Kumar. HamNoSys to SiGML Conversion System for Sign Language Automation. In *Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016)*, volume 89 of *Procedia Computer Science*, pages 794–803, 2016. doi: https://doi.org/10.1016/j.procs.2016.06.063.

Michael Kipp, Alexis Heloir, and Quan Nguyen. Sign language avatars: animation and comprehensibility. In Hannes Högni Vilhjálmsson, Stefan Kopp, Stacy Marsella, and Kristinn R. Thórisson, editors, *International Workshop on Intelligent Virtual Agents*, pages 113–126, Reykjavik, Iceland, 2011. Springer. doi: 10.1007/978-3-642-23974-8_13.

Philipp Koehn. Europarl: A parallel corpus for statistical machine translation. In *Proceedings of Machine Translation Summit X: Papers, MTSummit 2005, Phuket, Thailand, September 13-15, 2005*, pages 79–86, 2005. URL https://aclanthology.org/2005.mtsummit-papers.11.

Maria Kopf, Marc Schulder, and Thomas Hanke. Overview of datasets for the sign languages of Europe. Project deliverable D6.1, EASIER Consortium, 2021. URL https://doi.org/10.25592/UHHFDM.9560.

Maria Kopf, Marc Schulder, and Thomas Hanke. The Sign Language Dataset Compendium: Creating an overview of digital linguistic resources. In Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Julie A. Hochgesang, Jette Kristoffersen, Johanna Mesch, and Marc Schulder, editors, *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources*, pages 102–109, Marseille, France, June 2022a. European Language Resources Association (ELRA). ISBN 979-10-95546-86-3. URL https://www.sign-lang.uni-hamburg.de/lrec/pub/22025.pdf.

Maria Kopf, Marc Schulder, Thomas Hanke, and Sam Bigeard. Specification for the harmonization of sign language annotations. Project deliverable D6.2, EASIER Consortium, 2022b. URL https://doi.org/10.25592/UHHFDM.9842.

Paddy Ladd. *Understanding deaf culture: in search of deafhood.* Multilingual Matters, 2003. ISBN 9781853595479. doi: 10.21832/9781853595479.

Dongxu Li, Cristian Rodriguez, Xin Yu, and Hongdong Li. Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 1459–1469, 2020.

D. Lillo-Martin and J. Henner. Acquisition of sign languages. *Annual Review of Linguistics*, 7(1):395—419, 2021.

Colin Lualdi, Elaine Wright, Jack Hudson, Naomi Caselli, and Christiane Fellbaum. Implementing ASLNet V1.0: Progress and Plans. In Sonja Bosch, Christiane Fellbaum, Marissa Griesel, Alexandre Rademaker, and Piek Vossen, editors, *Proceedings of the 11th Global Wordnet Conference, GWC 2021, University of South Africa (UNISA), Potchefstroom, South Africa, January 18-21, 2021*, pages 63–72. Global Wordnet Association, 2021. URL https://aclanthology.org/2021.gwc-1.8/.

Richard P. Meier. Why different, why the same? Explaining effects and non-effects of modality upon linguistic structure in sign and speech. In Richard P. Meier, Kearsy Cormier, and David Quinto-Pozos, editors, *Modality and structure in signed and spoken languages*, pages 1–25. Cambridge University Press, 2002. doi: 10.1017/CBO9780511486777.001.

Johanna Mesch. Perspectives on the Concept and Definition of International Sign. In *Official programme of the 15th World Congress of the Deaf*. World Federation of the Deaf, 2010. URL http://wfdeaf.org/wp-content/uploads/2016/11/Perspectives-on-the-Concept-and-Definition-of-IS_Mesch-FINAL.pdf.

Johanna Mesch, Lars Wallin, and Thomas Björkstrand. Sign language resources in Sweden: Dictionary and corpus. In Onno Crasborn, Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Jette Kristoffersen, and Johanna Mesch, editors, *Proceedings of the LREC2012 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon*, pages 127–130, Istanbul, Turkey, 2012. European Language Resources Association (ELRA). URL https://www.sign-lang.uni-hamburg.de/lrec/pub/12001.pdf.

Hope E. Morgan, Onno Crasborn, Maria Kopf, Marc Schulder, and Thomas Hanke. Facilitating the spread of new sign language technologies across Europe. In Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Julie A. Hochgesang, Jette Kristoffersen, Johanna Mesch, and Marc Schulder, editors, *Proceedings of the LREC2022 10th workshop on the representation and processing of sign languages: Multilingual sign language resources*, pages 144–147, Marseille, France, 2022. European Language Resources Association (ELRA). ISBN 979-10-95546-86-3. URL https://www.sign-lang.uni-hamburg.de/lrec/pub/22026.pdf.

Amit Moryossef and Yoav Goldberg. Sign Language Processing, 2021. URL https://sign-language-processing.github.io/.

Amit Moryossef, Ioannis Tsochantaridis, Joe Dinn, Necati Cihan Camgöz, Richard Bowden, Tao Jiang, Annette Rios, Mathias Muller, and Sarah Ebling. Evaluating the immediate applicability of pose estimation for sign language recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3434–3440, 2021.

Medet Mukushev, Arman Sabyrov, Madina Sultanova, Vadim Kimmelman, and Anara Sandygulova. Towards semi-automatic sign language annotation tool: SLAN-tool. In Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Julie A. Hochgesang, Jette Kristoffersen, Johanna Mesch, and Marc Schulder, editors, *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources*, pages 159–164, Marseille, France, June 2022. European Language Resources Association (ELRA). ISBN 979-10-95546-86-3. URL https://www.sign-lang.uni-hamburg.de/lrec/pub/22030.pdf.

Irene Murtagh. *A Linguistically Motivated Computational Framework for Irish Sign Language*. PhD thesis, Dublin: Trinity College Dublin, 2019.

Irene Murtagh, Rachel Moiselle, and Lorraine Leeson. Sign Languages and Language Technology: Linguistic and Technical Challenges. In *Irish Association of Applied Linguistics (IRAAL) conference presentation*, 2021.

Irene Murtagh, Víctor Ubieto Nogales, and Josep Blat. Sign language machine translation and the sign language lexicon: A linguistically informed approach. In *Proceedings of the 15th biennial conference of the Association for Machine Translation in the Americas (Volume 1: Research Track)*, pages 240–251, Orlando, USA, September 2022. Association for Machine Translation in the Americas. URL https://aclanthology.org/2022.amta-research.18.

Carolina Neves, Luísa Coheur, and Hugo Nicolau. HamNoSyS2SiGML: Translating HamNoSys Into SiGML. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, page 6035–6039, 2020.

Adrián Núñez-Marcos, Olatz Perez de Viñaspre, and Gorka Labaka. A survey on sign language machine translation. *Expert Systems with Applications*, 213:118993, 2023. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2022.118993. URL https://www.sciencedirect.com/science/article/pii/S0957417422020115.

Maria Papadogiorgaki, Nikolaus Grammalidis, Dimitrios Tzovaras, and Michael G. Strintzis. Text-to-sign language synthesis tool. In *13th European Signal Processing Conference*, pages 1–4, Antalya, Turkey, 2005.

Magdalena Pasikowska-Schnass. Sign languages in the EU. Technical report, European Parliamentary Research Service, 2018. URL http://www.europarl.europa.eu/RegData/etudes/ATAG/2018/625196/EPRS_ATA(2018)625196_EN.pdf.

Susan Plann. *A Silent Minority: Deaf Education in Spain, 1550-1835*. Online access: California Digital Library UC Press E-Books Collection, 1982-2004. University of California Press, 1997. ISBN 9780520204713. URL https://books.google.be/books?id=udHE4CEe8vkC.

Siegmund Prillwitz, Thomas Hanke, Susanne König, Reiner Konrad, Gabriele Langer, and Arvid Schwarz. DGS Corpus project – development of a corpus based electronic dictionary German Sign Language / German. In Onno Crasborn, Eleni Efthimiou, Thomas Hanke, Ernst D. Thoutenhoofd, and Inge Zwitserlood, editors, *Proceedings of the LREC2008 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, pages 159–164, Marrakech, Morocco, June 2008. European Language Resources Association (ELRA). URL https://www.sign-lang.uni-hamburg.de/lrec/pub/08018.pdf.

Lorna C. Quandt, Athena Willis, Melody Schwenk, Kaitlyn Weeks, and Ruthie Ferster. Attitudes toward signing avatars vary depending on hearing status, age of signed language acquisition, and avatar type. *Frontiers in Psychology*, 13, 2022. ISSN 1664-1078. doi: 10.3389/fpsyg.2022.730917. URL https://www.frontiersin.org/articles/10.3389/fpsyg.2022.730917.

Razieh Rastgoo, Kourosh Kiani, and Sergio Escalera. Sign language recognition: A deep survey. *Expert Systems with Applications*, 164:113794, 2021. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2020.113794. URL https://www.sciencedirect.com/science/article/pii/S095741742030614X.

Ben Saunders, Necati Cihan Camgöz, and Richard Bowden. Progressive Transformers for End-to-End Sign Language Production. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020.

Ineke Schuurman, Thierry Declerck, Caro Brosens, Margot Janssens, Vincent Vandeghinste, and Bram Vanroy. Are there just wordnets or also signnets? In *Proceedings of the 12th International Global Wordnet Conference*, Donastia, Spain, 2023.

Umar Shoaib, Nadeem Ahmad, Paolo Prinetto, and Gabriele Tiotto. Integrating MultiWordNet with Italian Sign Language lexical resources. *Expert Systems with Applications*, 41(5):2300–2308, 2014. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2013.09.027.

Ozge Mercanoglu Sincan and Hacer Yalim Keles. Autsl: A large scale multi-modal turkish sign language dataset and baseline methods. *IEEE Access*, 8:181340–181355, 2020.

Stephanie Stoll, Necati Cihan Camgöz, Simon Hadfield, and Richard Bowden. Text2Sign: Towards Sign Language Production Using Neural Machine Translation and Generative Adversarial Networks. *International Journal on Computer Vision*, 128:891–908, 2020. URL https://doi.org/10.1007s11263-019-01281-2.

S. Augustine Su and Richard Keith Furuta. VRML-based representations of ASL fingerspelling on the World Wide Web. In *Proceedings of the third international ACM conference on Assistive technologies*, pages 43–45, 1998. doi: 10.1145/274497.274506.

Svenskt teckenspråkslexikon. Swedish sign language dictionary, 2023. URL https://teckensprakslexikon.su.se. Online dataset.

Alan M. Turing. Computing machinery and intelligence. *Mind*, LIX(236):433–460, 1950. ISSN 0026-4423. doi: 10.1093/mind/LIX.236.433. URL https://doi.org/10.1093/mind/LIX.236.433.

Britt Van Gemert, Richard Cokart, Lyke Esselink, Maartje De Meulder, Nienke Sijm, and Floris Roelofsen. First steps towards a signing avatar for railway travel announcements in the Netherlands. In *Proceedings of the 7th International Workshop on Sign Language Translation and Avatar Technology: The Junction of the Visual and the Textual: Challenges and Perspectives*, pages 109–116, Marseille, France, June 2022. European Language Resources Association. URL https://aclanthology.org/2022.sltat-1.17.

Myriam Vermeerbergen. *Grammaticale aspecten van de Vlaams-Belgische gebarentaal*. vzw Cultuur voor Doven, 1997. ISBN 9789075977035. URL https://books.google.be/books?id=J7XiAAAACAAJ.

Myriam Vermeerbergen. Past and current trends in sign language research. *Language & Communication*, 26(2):168–192, 2006. doi: 10.1016/j.langcom.2005.10.004.

Myriam Vermeerbergen and Mieke Van Herreweghe. *Wat geweest / gewenst is: organisaties van en voor doven in Vlaanderen bevraagd over 10 thema's*. Academia press, 2008.

Myriam Vermeerbergen, Lorraine Leeson, and Onno Crasborn. Simultaneity in Signed Languages: A String of Sequentially Organised Issues. In Myriam Vermeerbergen, Lorraine Leeson, and Onno Crasborn, editors, *Simultaneity in Signed Languages: Form and function*, pages 1–25. John Benjamins Publishing Company, 2007. doi: 10.1075/cilt.281.

Mark Wheatley and Annika Pabsch. *Sign Language Legislation in the European Union - Edition II*. European Union for the Deaf, Brussels, 2012. URL https://www.eud.eu/wp-content/uploads/2021/11/EUD_SL_II.pdf.

Beatrijs Wille. Assessing the visual communication and Flemish Sign language development in deaf two-year-olds. In *Collated Papers for the ALTE 7th International Conference*, 2021.

Peter Wittenburg, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes. ELAN: a professional framework for multimodality research. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy, May 2006. European Language Resources Association (ELRA). URL http://www.lrec-conf.org/proceedings/lrec2006/pdf/_pdf.pdf.

Bencie Woll, Neil Fox, and Kearsy Cormier. Segmentation of signs for research purposes: Comparing humans and machines. In Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Julie A. Hochgesang, Jette Kristoffersen, Johanna Mesch, and Marc Schulder, editors, *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources*, pages 198–201, Marseille, France, June 2022. European Language Resources Association (ELRA). ISBN 979-10-95546-86-3. URL https://www.sign-lang.uni-hamburg.de/lrec/pub/22007.pdf.

World Federation of the Deaf. Who we are, 2023. URL https://wfdeaf.org/who-we-are/.

Shicai Yu and Rong Lu. Research of Sign Language Synthesis Based on VRML. *Applied Mechanics and Materials*, 347–350:2631–2635, 2013. doi: 10.4028/www.scientific.net/AMM.347-350.2631.

Appendix

A. European Sign Languages

Table 1 lists all European countries which are members of the European Union for the Deaf. For each country the table lists which sign languages are used and how many signers there are.

Country	Sign Languages	Abbrev.	No. of signers
Austria	Österreichische Gebärdensprache	ÖGS	8,000
Belgium	Vlaamse Gebarentaal	VGT	5,000
	Langue des Signes de Belgique Francophone	LSFB	5,000
Bulgaria	Български жестомимичен език (Bŭlgarski zhestomimichen ezik)	BŽE	5,000
Croatia	Hrvatski znakovni jezik	HZJ	6,500
Cyprus	Κυπριακή Νοηματική Γλώσσα	ΚΝΓ	1,000
Czech Republic	Český Znakový Jazyk	CSJ	10,000
Denmark	Dansk tegnsprog	–	4,000
Estonia	Eesti viipekeel	EVK	1,500[36]
	Russian Sign Language		
Finland	Suomalainen viittomakieli	SVK	3,000
	Finlandssvenskt teckenspråk	SRVK	90
France	Langues des Signes Française	LSF	120,000
Germany	Deutsche Gerbärdensprache	DGS	83,000
Greece	Ελληνική Νοηματική Γλώσσα	ΕΝΓ	5,000
Hungary	Magyar jelnyelv	–	9,000
Iceland	Islenskt táknmál	–	250
Ireland	Teanga Chomharthaíochta na hÉireann / Irish Sign Language	ISL	5,000
Italy	Lingua dei Segni Italiana	LIS	40,000
Latvia	Latviešu Zīmju Valoda	LZV	2,000
Lithuania	Lietuviu gestu kalba	LGK	8,000
Luxemburg	Deutsche Gebärdensprache	DGS	250
Malta	Lingwa tas-Sinjali Maltija	LSM	200
The Netherlands	Nederlandse Gebarentaal	NGT	7,500
Norway	Norsk Tegnspråk	NTS	5,500
Poland	Polski Jezyk Migowy	PJM	50,000
Portugal	Lingua Gestual Portuguesa	LGP	60,000
Romania	Limbaj Mimico-Gestual Romanesc	LMGR	24,601
Slovakia	Slovnik Posunkovej Reci	–	5,500
Slovenia	Slovenski znakovni jezik	–	1,021
Spain	Lengua de Signos Espanola	LSE	100,000[37]
	Lengua de Signos Catalna	LSC	
Sweden	Svenskt Teckenspråk	SSL	8,000
Switzerland	Deutschschweizer Gebärdensprache	DSGS	6,750
	Langue des Signes Francaise	LSF	2,750
	Lingua dei Segni Italiana	LIS	500
United Kingdom	British Sign Language	BSL	87,000[38]
	Irish Sign Language	ISL	

Table 1: European countries with their respective number of sign language users

[36] EVK and RSL together
[37] LSE and LSC together
[38] BSL and ISL together