DOI: 10.17469/O2106AISV000002

CARLO GERACI, ROLAND PFAU, PIETRO BRAIONE, CARLO CECCHETTO, JOSEP QUER

Hidden languages in a digital world: The case of sign language archives

SIGN-HUB is a European project involving collaborators from seven countries funded within the Horizon 2020 Research and Innovation program. The scope of the project is both socio-cultural and linguistic, as its aim is to document and preserve the culture, the history, and the languages of European Deaf communities. After a brief description of the various components of the project, we focus on the documentation of the life stories of Deaf people and the creation of a digital sign language archive.

Key words: sign language, digital archive, Deaf culture, documentation, preservation.

1. *Introduction*

Sign languages are the natural languages used by Deaf[1] people in everyday communication. They have been shown to display complex grammatical structures fully on a par with spoken languages, despite the fact that they employ a different modality for signal production and perception (Sandler, Lillo-Martin, 2006; Pfau, Steinbach & Woll, 2012). That is to say, while spoken languages are based on the auditory-vocal modality, sign languages make use of the visual-spatial modality. Most sign languages are minority languages in the sense that they are typically immersed in an environment with a dominant spoken language[2], and indeed they share some features with minority spoken languages: they are often marginalized or even discriminated, they are not taught at schools, and their users may constitute a subculture within the mainstream culture (see, e.g., Burns, 1996 for Irish Sign Language and Mougeon, Nadasdi, 1998 for a general perspective on minority language communities). However, they also differ from minority spoken languages in several ways: beside the modality of transmission, the pattern of language acquisition is unique, since typically only 5-10% of deaf children are born into Deaf families where sign language is present, and most Deaf signers are exposed to sign language beyond the very first years of life (see Quer, Steinbach, 2019).

[1] We follow the by now well-established convention of distinguishing 'deaf' as the physical condition of lack of hearing (with lowercase d) from 'Deaf' as the cultural and linguistic identity of signers (with capital D).

[2] A notable exception are shared sign languages, sometimes also called rural or village sign languages, a term that identifies sign languages that are used in small communities with an unusually high incidence of (often hereditary) deafness. In such contexts, deafness is often less stigmatized, and a high percentage of hearing community members is also fluent in the local sign language. For an overview of shared sign languages and their specific linguistic properties, see Nyst (2012) and De Vos, Pfau (2015).

One peculiarity of sign languages is that none of the Deaf communities of the world has independently developed a writing system for their sign language, the main reason being that Deaf people are generally educated in the dominant spoken language (i.e., they can at least read and write the dominant spoken language, albeit at varying levels of proficiency)[3]. The direct consequence of this fact is that sign language communities are prominently "oral" communities (Byrne, 2016), which implies that their culture – including artistic expressions like storytelling, poetry, and theatre – is transmitted "orally" (i.e. in a non-written, visual form) via sign language and has therefore, for the longest time, not been documented[4].

In fact, the lack of a writing system or other suitable ways to transmit cultural productions is one of the main reasons why historians of Deaf communities always have troubles in finding first-hand documents. Sign language as a means of communication for the Deaf is already mentioned in Plato's *Cratilus*, but the oldest film document of a signer probably comes from American Sign Language (ASL) and dates back only to 1913. In this film, George Veditz, the president of the National Association of the Deaf, talks about the *Preservation of the Sign Language* (film title) and famously expresses "As long as we have deaf people on earth, we will have signs". This and subsequent movies are now considered some of the most significant documents in Deaf history (for historical ASL materials, see Supalla, 2001 and Supalla, Clark, 2015)[5].

In this paper, we describe the preliminary results of the Sign-Hub project, the first systematic attempt (i) to document the history, the culture, the experiences, the identity, and the languages of various European Deaf communities; and (ii) to store all of this information and make it available via a digital on-line platform. That is, the project has two closely related goals, which, in a sense, are like two sides of a medal: on the one hand, it is devoted to the creation of sign language-related content; on the other hand, it involves the development of a digital infrastructure that a) allows for immediate access to sign language content, and b) guarantees long term preservation of the digital files.

In the remainder of the paper, we first offer a general overview of the Sign-Hub project (Section 2). In Section 3, we then offer a description of the components of the sign language archive. The solutions we are implementing for sharing, preserving, and protecting the digital files are described in Section 4. Section 5 concludes the paper and addresses some of the future challenges.

---

[3] There are various transcription systems that employ specific symbols for representing the sub-lexical components of signs. However, these systems are highly intricate and are only used for scientific purposes (for an overview, see Frishberg, Hoiting & Slobin, 2012).

[4] Of course, there are many books and papers – both of a scientific and a more popular nature – on sign languages, Deaf culture, Deaf communities, and Deaf education. But crucially, these publications are not composed in the language of the population they talk about. For a detailed description of the Deaf world, see Bauman (2008), Gertz, Boudreault (2016), Goodstein (2006), and Parasnis (1996). Quotes from Deaf subjects, for instance, are usually translated into the language of the publication. See for instance, Padden, Humphries (2005) and Leigh, Andrews & Harris (2018).

[5] The video can be viewed on Youtube.

2. *The SIGN-HUB project*

SIGN-HUB (2016–2020) is a European project funded within the Horizon 2020 Research and Innovation program[6]. It involves a network of ten universities and research centers from seven different countries, coordinated by Josep Quer from the Universitat Pompeu Fabra in Barcelona[7]:

- France: CNRS, Institut Jean-Nicod, Paris – coordinator: Carlo Geraci;
  University of Paris-Diderot – coordinator: Caterina Donati;
- Germany: Georg-August University Göttingen – coordinator: Markus Steinbach;
- Israel: University of Tel Aviv – coordinator: Naama Friedmann;
- Italy: Ca' Foscari University, Venice – coordinators: Chiara Branchini and Anna Cardinaletti;
  CINI, Milan – coordinator: Mauro Pezze;
  University of Milano-Bicocca – coordinator: Carlo Cecchetto;
- The Netherlands: University of Amsterdam – coordinator: Roland Pfau;
- Spain: Universitat Pompeu Fabra, Barcelona – coordinator: Josep Quer;
- Turkey: Bogaçizi University, Istanbul – coordinator: Meltem Kelepir.

Within the general frame of preserving the culture, the history, the identity, and the languages of European Deaf communities, SIGN-HUB has two main goals: the production of materials and contents related to Deaf communities and their sign languages and the development of the digital infrastructure needed to make the content accessible, and to protect and preserve it in the long-term.

2.1 Documenting sign languages and Deaf communities

There are four main outputs of this part of the project: (i) grammars of six European sign languages, (ii) a sign language atlas, (iii) clinical tests for the assessment of sign language abilities in special populations, and (iv) an archive of the life stories of elderly Deaf signers. Here, we briefly address the first three components, while the last one will be discussed in more detail in Section 3.

The SignGram Blueprint is an open access tool for language specialists and sign language linguists that provides detailed information on how to write a sign language grammar (Quer, Cecchetto, Donati, Geraci, Kelepir, Pfau & Steinbach, 2017)[8]. In itself it is not a grammar, but it contains (i) a checklist of relevant grammatical phenomena and processes from which the table of content of a grammar can be built; (ii) a detailed discussion of background information related to the linguistic features, phenomena, and constructions listed in the checklist; these are

[6] The SIGN-HUB grant agreement number is 693349.

[7] More information about the teams and the specific goals of each research unit can be found on the project website (https://www.unive.it/pag/33750/) and on the national SIGN-HUB websites.

[8] This publication is the outcome of another European project, SignGram (2011-2015), of which all the participants to SIGN-HUB (except Israel) were members (COST Action 1006: Unravelling the grammars of European Sign Languages; European Science Foundation). The SignGram Blueprint is freely accessible at: https://www.degruyter.com/viewbooktoc/product/467598.

described from the perspective of languages in the visual-gestural modality; and (iii) guidelines on how to identify and analyze linguistic phenomena (including suggestions for data collection) as well as relevant references. On the basis of these detailed guidelines, grammars of the following six sign languages will be provided within SIGN-HUB: Catalan Sign Language (LSC), German Sign Language (DGS), Italian Sign Language (LIS), Sign Language of the Netherlands (NGT), Spanish Sign Language (LSE), and Turkish Sign Language (TİD). One crucial aspect of these grammars is the fact that, in addition to the usual text descriptions, various types of visual materials (images and digital videos) will be included. Besides the intrinsic value of detailed language descriptions, this is the first step towards the long-term preservation of the individual sign languages. We expect the grammars of these languages to be used by language instructors and teachers as a tool for creating additional teaching materials for sign language courses. We also expect that the grammars will give increased visibility – and thus status – to each of these languages, and that local Deaf communities may capitalize on such a resource to advocate in favor of sign language recognition where it has not happened yet, and/or new policies in the school education of Deaf children (e.g., bilingual-bimodal educational approaches).

Along with in-depth language descriptions, SIGN-HUB also aims at describing typological variation across sign languages (e.g., Perniss, Pfau & Steinbach, 2007; Zeshan, Palfreyman, 2017). This goal will be achieved by developing an atlas dedicated to sign language structures, modelled on existing ones like WALS, SSWL, and APiCS[9]. Specifically, four questionnaires have been prepared (phonology & lexicon, morphology, syntax, pragmatics & socio-history). Each questionnaire is designed to include around fifty questions targeting a variety of aspects from each domain of sign language linguistics. The plan is to collect information from at least one-hundred sign languages from all over the world. The questionnaires will be filled in by experts in the particular sign language, and, if available, additional experts from different domains can be consulted for providing missing information for a particular language. We expect the sign language atlas to have a large and immediate impact on the study of typological variation across sign languages. Given the design of the questionnaires and their implementation into the on-line atlas, we expect the data collection of this part of the project to continue after 2020 with new questionnaires covering additional sign languages to be added in the future. As with the sign language grammars, in this respect, too, local Deaf communities can widely benefit from this tool. In smaller communities in particular, realizing that grammatical phenomena or features attested in their sign language are also found in other (related or unrelated) languages will likely reinforce (or even initiate) the process of language awareness, upon which Deaf identity is built.

Accurate detection of language disorders is a key factor for early intervention in clinical linguistics, both for adults with post-stroke or progressive aphasia and for

[9] More information about these projects can be found at the following websites: WALS (https://wals.info), SSWL (http://test.terraling.com/groups/7), and APiCS (https://apics-online.info).

children with developmental language disorders. Significant progress has been made on spoken language assessment tools. However, nothing comparable has been put in place for sign languages (see Mann, Haug, 2014 for an overview of assessment tests in sign language), although valuable work has been done for some sign languages, i.e., ASL (Hauser, Paludneviciene, Riddle, Kurz, Emmorey & Contreras, 2016) and British Sign Language (for more info see DCAL Assessment Portal, https://dcal-portal.org). The third component of SIGN-HUB is therefore dedicated to creating a variety of assessment tests to detect lexical impairment and measure proficiency in various grammatical domains. One key aspect of the project is to standardize the tests across healthy populations of signers with various levels of sign language exposure. Indeed, the signing population is quite stratified and includes Deaf native signers (typically, deaf people born into signing families), who represent only a small percentage of deaf signers, signers with early exposure to sign language (deaf people born into hearing families who are exposed to sign language before school age), and late learners (deaf people who are exposed to sign language only during or after the school age). While hardly detectable in normal conversation (at least in some cases), the various groups of signers may have different levels of competence with respect to various, often subtle, properties of the grammar (for recent discussion, see Mayberry, Kluender, 2018). The tests will cover all the areas of SL grammar including phonology, morphology, lexicon, syntax and semantics. We expect the tests to be systematically used for early diagnosis and clear identification of language pathologies both in Deaf adults and children.

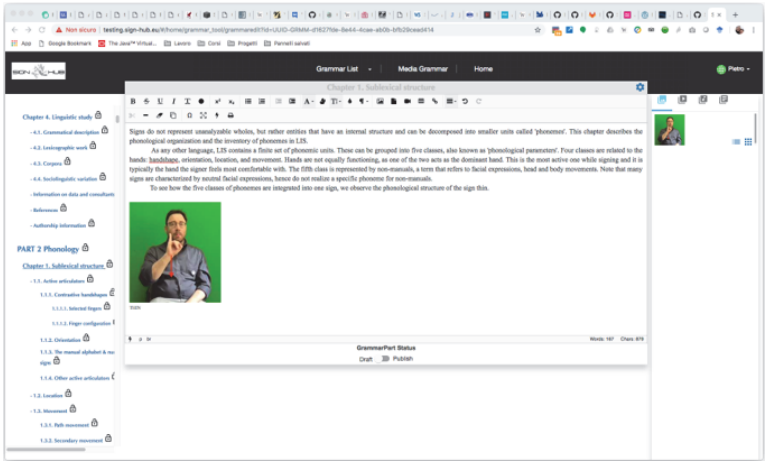
2.2 Digital infrastructure

This part of the project is dedicated to the creation of an on-line digital platform to host the materials and products described in Section 2.1. Without entering into the technical details of the coding, there are two parts of the platform with which people will be able to interact. One is the set of end-user interfaces/tools developed to implement each of the four research components of SIGN-HUB (grammars, atlas, assessment tests, digital archive); the other is the set of end-user interfaces that will allow display of and access to the final products of each research component. At this stage of the project, only the first set of tools has been developed, which are the ones used by SIGN-HUB members to upload and create their materials. The tools for the grammar writers, atlas questionnaires, and language assessment are briefly described here. Issues related to the video archive will be addressed in Section 3.3.

The grammar tool is designed for grammar writers to type text materials (including tables) directly onto the platform and/or to upload their materials (e.g., texts produced with external software, images, and videos). As illustrated in Figure 1, the workspace is divided into three main areas. On the left side, the grammar writer finds the table of content, which is based on the SignGram checklist, and which allows for easy navigation from one part of the grammar to another by clicking on the links. The central part contains a rich text editor, where text can either be typed or be pasted already formatted from other sources. The right side contains

a quick access to the database of visual materials (images and videos) to be used in the grammar.

Figure 1 - *Outline of the Grammar Tool*



Figures 2 and 3 illustrate the workspace of the atlas and the assessment components. The basic concept is that the questions of a questionnaire (atlas) and the items of a test (assessment) are structured like presentation slides. The buttons on the right side of the workspace offer quick access to direct measures (e.g., multiple answers, mutually exclusive checkboxes, continuous scales, etc.), while the central area corresponds to the slide itself. An adjustable grid allows for adapting the positioning of the various objects, text boxes, clickable images, videos, etc. Finally, on the right side, visual materials from the SIGN-HUB database of images and videos can be quickly dragged and dropped on the slide.

Figure 2 - *Outline of the Atlas Tool*

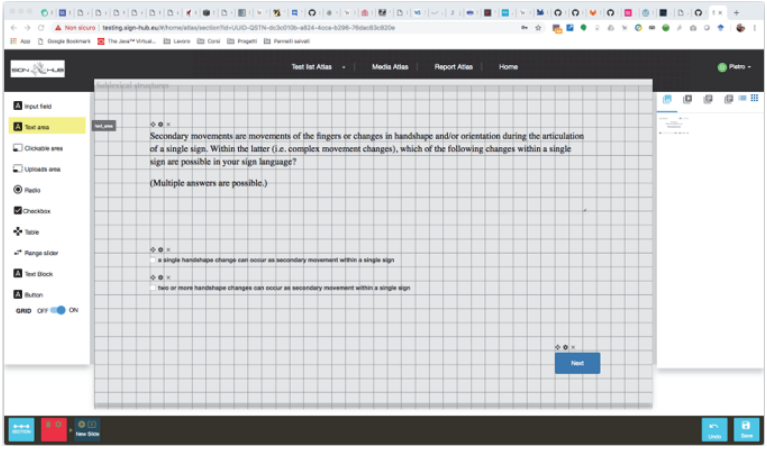
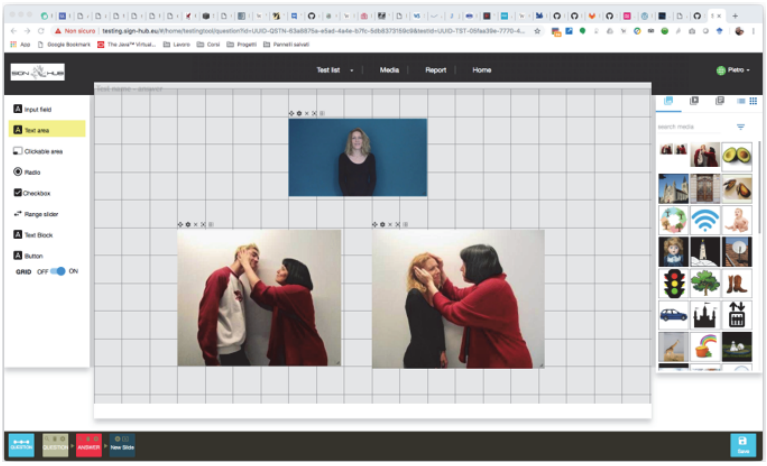


Figure 3 - *Outline of the Assessment Tool*



3. *Life stories of Deaf people*

3.1 Significance and urgency

As already mentioned in the introduction, Deaf communities have not developed writing systems for sign languages for everyday use. Therefore, Deaf communities are intrinsically bound to "oral" transmission of their cultures in the sense of passing on cultural values, traditions, and productions in a non-written form. One clear example is the role of narratives in the development of a Deaf identity in deaf children (Sutton-Spence, 2010)[10]. Although numerous publications document (aspects of) Deaf communities and their languages (e.g., Monaghan, Schmaling, Nakamura & Turner, 2003; Woll, Ladd, 2003; McCaskill, Lucas, Bayley & Hill, 2011), access to primary data has always been problematic until video recording tools have been made available at reasonable prices. Indeed, it was only after cameras and videotapes have been accessible to "non-professionals" that cultural events of Deaf people started being documented. There are three crucial facts which make starting a systematic documentation of Deaf history, culture and communities particularly urgent and compelling.

First, the first half of the 20th Century has witnessed some of the most devastating events of human history (the two World Wars, the Shoah, the atomic bomb). Of course, Deaf people have their own perspective on these (and other) events. For instance, most Deaf children were left in residential schools during World War II with little information about what was happening around them and why (e.g., bombing, starvation, sterilization programs, etc.), as explained in the documentary *1939–1945 Que Faisaient Les Sourds?* ('1939–1945 What did the Deaf do?') on the experiences of Deaf people in France during the second world war. Yet, given

[10] See also Rutherford (1993) and Rayman (1999); for more comprehensive overviews of aspects of Deaf Culture (in the United States), see Padden, Humphries (2005) and Leigh, Andrews & Harris (2018).

that the survivors are now elderly signers, we are left with a very short time-window for documenting their unique memories.

Second, the entire approach to deaf education has radically changed in recent years, in part due to the need of social integration between hearing and deaf children, in part due to the fact that (hearing) parents of deaf children now commonly choose for a cochlear implant (Spencer, Marschark, 2003; also see Humphries, Kushalnagar, Mathur, Napoli, Rathmann & Smith, 2019 for an overview of strategies to support parents with deaf children). As a result, (residential) deaf schools disappeared almost everywhere, and deaf children are nowadays mostly educated in mainstream schools with some additional provisions (e.g., interpreters, special educators, communication assistants, etc.). This apparently harmless shift has radically changed access to sign language. In the past, even in strictly oralist schools, where sign language was forbidden in class, deaf pupils often used sign language to communicate with each other outside the classroom. Now deaf children can be scattered across mainstream schools, with few (sometimes only a single one) in an entire school. If this happens, the common way for deaf children to have spontaneous access to sign language from peers is basically eliminated (but see Marschark, Tang & Knoors, 2014). In some countries, this change in education happened quite suddenly in the mid Seventies, inducing a significant gap between older and younger generations of signers. Third, even for the rare cases in which some memories have been collected on video, the paradigm changes imposed by the digital era make these recordings quickly obsolete and useless (Castells, 1996).

Taken together, the memories of elderly Deaf signers are a treasure that is at risk of being lost forever (also see Legg, 2016). In the next sections, we will illustrate the steps that have been taken within SIGN-HUB to preserve this treasure. We adopted two strategies: in some of the participating countries, we conducted semi-structured interviews; in others, we digitized documentation that was already available. We briefly describe our methodology of data collection/selection in the next section, and then we summarize the main findings in Section 3.3.

3.2 Materials

A crucial component of our documentation and preservation effort are the interviews with Deaf elderly signers conducted in five of the participating countries, Germany, Italy, the Netherlands, Turkey and Spain[11]. Our goal has been to interview at least 20 signers above the age of 65 per country. In the end, we conducted between 21 and 25 interviews per country, the age range of interviewees being between 66 and 93 years.

All interviews were conducted by Deaf research assistants fluent in the local sign language. They vary considerably in length, but all of them are semi-structured in that they follow a questionnaire that has been developed specifically for the interviews. The questionnaire is organized around various topics (e.g., family, Deaf experiences, historical events), and interviewers were encouraged to cover each of the

[11] Interviews were conducted in six different sign languages, as two sign languages were involved in Spain, i.e., LSC and LSE.

topics. Still, it was also made clear at the outset that the questionnaire was meant as a guideline. Interviewers were informed that they did not have to strictly follow the order of questions, and that the inclusion of other topics that came up during the interview was welcome. The most important consideration was to create a comfortable and safe atmosphere for the interviewee. During basically all interviews, it became clear that the elderly Deaf signers enjoyed sharing their life stories. In fact, they would often bring photos and other memorabilia to share with the interviewer. In total, 137 interviews were conducted of a total length of approximately 175 hours.

All interviews were filmed with two cameras: one on both the interviewer and the interviewee, and one on only the interviewee. Participants were informed about the purpose of the interviews and signed a consent form. Although the video recordings are meant to preserve memory and may at times include non-frontal shooting or the interviewee leaving the scene for a moment, the quality of the videos is such that they can be reused for other purposes. Besides their obvious historical and cultural value, the interviews are also of linguistic use in that they contribute to the documentation and preservation of SLs. The possibility of subjecting the productions of elderly signers to linguistic analysis will thus allow a direct comparison with the sign languages used by younger generations, as commonly done in more traditional corpus-based sociolinguistic studies where signers' age is used as a predictor of language variability (Lucas, Bayley & Valli, 2001; Geraci, Battaglia, Cardinaletti, Cecchetto, Donati, Giudice & Mereghetti, 2011; Fenlon, Schembri, Johnston & Cormier, 2015). Such comparative sociolinguistic research has been initiated within SIGN-HUB by annotating fragments of the videos with the computer-based annotation tool ELAN (Crasborn, Sloetjes, 2010). Finally, varying amounts of data per country have also been subtitled in the local spoken language and English in order to be shared with international non-signing audiences (for instance, in the context of a documentary movie that has been created).

Besides the interviews, we will also include in the digital archive materials that exist outside the SIGN-HUB project. Certain aspects of the life of Deaf people in the 20th century are relatively well documented for the Deaf communities of France and Israel. These existing materials will also constitute an important part of our archive.

As for France, the CNRS video library already contains an English subtitled documentary from 2000 entitled "Deaf witnesses, silent witnesses", directed by Brigitte Lemaine and Stéphane Gatti. The documentary, which is publicly available[12], describes the consequences on the Deaf community of the racial hygiene law imposed by Hitler in 1933, which lead to the sterilization of many people with disabilities including Deaf people; it also addresses obligatory abortions and euthanasia in the context of the so-called T4 program. In addition to this, the Académie de la Langue des Signes Française, a Deaf association based in Paris, has a rich archive of videotapes documenting a variety of activities and events that happened in the

[12] The streaming of the documentary is available at the following address: https://videotheque.cnrs.fr/doc=905.

second half of the 20th century. Some of these videotapes are unique copies already in a deteriorated state. The president of the association, Ronit Leven, selected three movies that are most important to preserve, and these have already been digitized (for the procedure see Section 4). All three are amateur documentaries shot by people working at the Académie de la Langue des Signes Française:

– *1939–1945 Que Faisaient Les Sourds?* ('1939–1945 What did the Deaf do?')
  The movie reports on an event organized by the Académie on March 19, 2004, which dealt with the situation of Deaf people during the Second World War. Footage of the live event is intertwined with short clips on WWII and interviews with both Deaf and hearing signers.
– *La Vie Des Sourds Pieds Noirs Et Juifs D'Algérie* ('Life of the Deaf *Pieds Noirs* and Jews of Algeria')
  This movie reports on a similar event (date unknown), this time about the war in Algeria (1954–1962).
– *Quel Avenir Pour Les Personnes Agées Sourdes?* ('What is the future of the elderly Deaf?')
  This is a collection of interviews with elderly Deaf people in France conducted in the early 2000s.

The digitized documentaries are now in the process of being subtitled in French. Subsequently, they will be voiced-over in French, and, in a third step, English subtitling will also be provided. A noteworthy aspect that makes these movies a unique and crucial contribution to the documentation of Deaf history is the fact that they have been realized either entirely by Deaf people or under the direct supervision and direction of Deaf people in an effort to preserve their own memories. They therefore represent an impressive symbol of Deaf identity and of a minority community.

Turning now to Israel, our original intention had been to digitize and include in the digital archive two types of documents[13]: (i) narratives from Deaf Holocaust survivors stored at the Yad Vashem World Holocaust Remembrance Center in Jerusalem, and (ii) life stories of Deaf elderly signers that had been recorded prior to our project. In the end, however, Yad Vashem decided against handing over their narratives to our open access digital archive. The life stories come from the Deaf archive of the University of Haifa, an impressive collection compiled by the late Irit Meir over the past 20 years. There are approximately 20 stories by elderly signers, but for budget reasons, only 4–5 of these will be included in the digital archive. The signers in the selected recordings are representative of different backgrounds in Israel, and the content of their stories is varied (e.g., Holocaust, immigration, experiences at Deaf school). Together, the stories have a length of roughly one hour, and they will be subtitled in Hebrew and English.

3.3 A glimpse at the life stories

Browsing the available materials – the interviews conducted in the context of the project as well as the pre-existing materials – it immediately becomes obvious that

[13] We are indebted to Rose Stamp for providing this information.

they include a wealth of invaluable information and anecdotes. As for the life stories, most of the interviewees readily shared both confronting and funny experiences. Although the analysis of the interviews, narratives, and digitized materials is still on-going, several recurrent topics are emerging across Deaf communities.

First, many signers talk about experiences during war time, including events that they, as Deaf subjects, may have experienced differently. As mentioned above, the Israel materials include the story of a Holocaust survivor. Second, many interviews include renditions of school experiences. This may be delightful anecdotes of the life at boarding schools for the deaf, but there are also shocking stories about neglect and abuse. Third, almost all signers share experiences of suppression and discrimination, e.g., at school or at the work place. This also includes discrimination against the use of sign language, which leads to the common experience of being excluded. Fortunately, at the same time, the signers also share stories about Deaf empowerment and increasing independence of Deaf people. Fourth, identity issues figure prominently in the narratives. These may have to do with the role of a Deaf person within a hearing family, but issues may also arise when a Deaf person identifies with multiple minorities (e.g., Deaf immigrants or Deaf homosexual people). Finally, some fragments are interesting from a linguistic perspective, as they reflect metalinguistic awareness, e.g., concerning lexical change or the origin of name signs.

The multifarious content of the narratives will be reflected in two outcomes of the SIGN-HUB project: a 40-minute documentary movie and an edited volume. The guiding topic of the movie will be 'independence (or lack thereof)'. The edited volume is expected to contain 13 chapters on cultural, socio-historical, and linguistic aspects of Deaf elderly signers. Most SIGN-HUB countries will contribute; in addition, a number of scholars external to SIGN-HUB (from Belgium, Germany, the Netherlands, and the US) have agreed to contribute a chapter related to the topic of the volume.

In addition, the three French documentaries, all complete interviews (including those that will not make it into the documentary), and the complete narratives from Israel will be part of the SIGN-HUB digital archive.

4. *The digital archive*

The easiness with which people can take videos from digital devices (computers, phones, tablets, etc.) and the storage systems that, as private users, virtually everybody can have access to (e.g., hard-drives, personal cloud space, etc.) may lead people to thinking that digital files are always accessible and imperishable. Of course, this is not the case either because physical space on individual devices is limited or because several factors may undermine access to digital files in the long term (e.g., evolving formats, upgrading software, removal of support by software developers, etc.). Notice that, hardware storage resources like DVDs, CD-ROMs are quickly becoming obsolete both for immediate access and long-term preservation uses. Furthermore, these supports are not ideal once the estimated amount/size of the

video materials that SIGN-HUB will produce is considered. Luckily, the problem of digital space for academic projects has been tackled at a higher level. In fact, a few platforms at the European level are now offering digital space to host the products of academic research projects, like DARIAH (https://www.dariah.eu) and CLARIN (https://www.clarin.eu), to mention just two of them.

In this section, we describe the main points of the SIGN-HUB data management plan, with a particular focus on the procedure adopted to digitize, store and make accessible the video materials produced within SIGN-HUB.

4.1 Digitizing old videos and collecting new materials

One of the most complex challenges that the teams working within SIGN-HUB has been faced with is to cope with the need of having high quality videos, the limited storage resources for digital materials, the varieties of standards and devices used in each research unit and the status of already collected materials. Overall there are two approaches: a top-down and a bottom-up approach. In the former approach, a uniform project-internal policy is used for all units imposing the use of similar (if not identical) hardware materials for filming, similar environmental conditions during the recordings, exporting in a pre-defined format, etc. In the latter approach, almost complete flexibility is given to individual units (e.g., mild requirements on the formatting could be given).

Within SIGN-HUB, we adopted a mixed approach depending on the requirements of specific parts of the project. For instance, the materials to be used as part of the assessment tests have to be of the best quality possible and must be as uniform as possible both within a research unit (i.e., the various tests that each units builds must have similar quality), and across units (i.e., the same standard should apply across the research units to make data comparable). On the other hand, for the interviews, the only requirement is that files are saved/exported in MPEG 4 format (H.264, 30 fps). The same requirement applies to the video materials already collected from previous projects. The estimated size of all video materials expected for the project is above 6 TB. Table 1 reports the amount of video materials divided by each country.

Table 1 - *Estimated size of video materials*

Sign language	Estimated total size of produced video files
LIS	3 TB
LSC & LSE	600 GB
NGT	1 TB
TİD	1 TB
LSF	70 GB
DGS, ISL	unknown
Total	>6 TB

One particular challenge for the project was the digitization of the three videotapes shared with the project by the Académie de la Langue des Signes Française. Indeed,

the technology for converting tapes is already obsolete and very few centers have it. The Research Center of the Italian public broadcasting television, Centro Ricerche RAI, offered its facility for the conversion[14].

The analog to digital conversion process involved three machines, a video cassette recorder for playback of the video (JVC BR-6400TR), a scan converter providing time base correction (SONY DSC-1024 G) and an analog-to-digital converter (DataVideo Y.U.V to DV DAC-2). Adobe Premiere CS6 was then used for video capturing and editing. The files are then exported in RAW format and in mp4 compressed format.

4.2 Immediate access vs. long-term storage

The ultimate goal of SIGN-HUB is the preservation of culture, identity, memories and languages of European Deaf communities, by giving the largest possible audience access to sign language grammars, atlas, documentaries and to have diagnostic tools to detect language disorders. The idea of implementing a centralized platform for a direct access to these contents seems to be the easiest way to achieve that goal. The activities described in Sections 2 and 3 are a big step towards that goal; however, on the technological side preserving digital files and giving access to them are two almost completely separate goals. Simplifying the issue, the expectation of generalized access to content can be exemplified by a well-organized website with a search engine that provides the requested content directly on the web browser, while the expectation of long-term preservation is that the quality of video remains intact across technological evolution of formats, devices, software, etc. The former expectation is generally met by allowing video files to be always accessible by streaming via internet. The latter expectation is met once data integrity and file fixity are maintained and metadata are stored in a standardized format. Once archived, the data are not immediately accessible and the process of retrieving data from the archive may take a considerable amount of time (which depends on the size of the file, the specific technique used to store the data, etc.).

Institutional providers that offer both services are quite rare. For instance, DARIAH only allows for long-term preservation, but not for streaming files, while only some centers of the CLARIN consortium offer both services, but the requirements for streaming are quite stringent. SIGN-HUB will use two French-based institutional infrastructures: HUMA-NUM (https://www.huma-num.fr) and ORTOLANG (https://www.ortolang.fr). HUMA-NUM is a large digital infrastructure that can be used for storage for all human sciences at the National and European level. ORTOLANG is a parallel network infrastructure which includes a repository of language data (corpora, lexicons, dictionaries, etc.) and readily available, well-documented tools for its processing specific for language and linguistic projects.

HUMA-NUM will host in its repository all files for long-term preservation (i.e., all data that need to be archived and not immediately available for streaming).

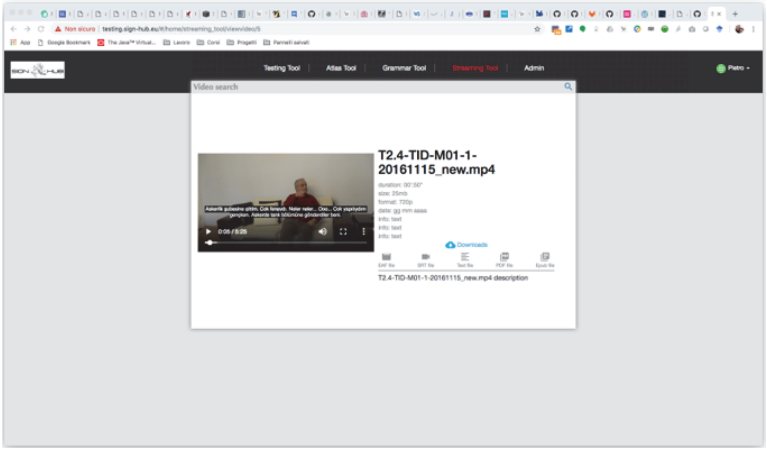
[14] We are very grateful to Andrea Del Principe for the help he provided.

ORTOLANG will host all video data that will be played in streaming, including e.g., video examples of the grammars, of the atlas questionnaires and final products, testing items for assessment, documentaries, etc. ORTOLANG will also host, on a virtual machine, all non-video materials, including the software that will be used to run the SIGN-HUB platform itself, the webpages, the databases of the assessment tests, the atlas questionnaires and the sign language grammars.

One clear advantage of using institutional repositories offering services at the European level, rather than private third-party services, is that they are by law compliant with security measures in terms of research data and sensitive data protection regulations. This is a delicate aspect of SIGN-HUB that cannot be fully addressed in this paper. However, to give a very brief outline of the challenges the project was faced with, consider that all data coming from healthy Deaf signers must be protected with the maximum level of security established by the European regulations. The reason is that, even when the purpose of these data is not medical at all, they come from a minority population with sensory disability. In this respect, institutional repositories offer a safety net in terms of data security.

The current layout of the streaming tool is illustrated in Figure 4. After selecting the video from the archive, users will have access to basic functions like full-screen size, subtitles and other options including downloading ELAN annotated files if available.

Figure 4 - *Layout of the streaming tool*



5. *Conclusions and perspectives*

In this paper, we presented the main goals of SIGN-HUB, an EU funded project whose main goal is to preserve the culture, the identity, the memories and the languages of European Deaf communities. In its core, SIGN-HUB contains a digital video-archive of Deaf communities, which are linguistic minorities using sign languages.

There are several directions in which SIGN-HUB can evolve after the end of the project. As far as the digital archive and the memories of (European) Deaf

communities are concerned, future technological developments will further reduce costs both for long-term storage and streaming (simply think about the apparently unlimited space offered for free by private video developers like YouTube, Vimeo, etc.). This opportunity will guarantee sustainability of the current project and preservation of the digital materials created during SIGN-HUB. In parallel, we hope to achieve quite a large impact on other signing communities both within and outside Europe, so that more sign language grammars will be available, the atlas will include more languages and more interviews of elderly signers are made available. In particular, we hope that non-urban signing communities will have access to our platform, on the one hand to benefit from our infrastructure; on the other hand, to extend the domain of sign language research. Another direction would be that of expanding the domain of the archive by including recordings of artistic performances like theater and poetry performances, which constitute an important component of Deaf culture in Western societies.

Bibliography

Bauman, H.-D.L. (Ed.) (2008). *Open your eyes: Deaf studies talking.* Minneapolis: University of Minnesota Press.

Byrne, A. (2016). Sign language literature. In Gertz, G., Boudreault, P. (Eds.), *The SAGE Deaf studies encyclopedia*. London: SAGE Publishing, 832-835.

Burns, S.E. (1998). Irish Sign Language: Ireland's second minority language. In Lucas, C. (Ed.), *Pinky extension and eye gaze: Language use in Deaf communities*. Washington, DC: Gallaudet University Press, 233-273.

Castells, M. (1996). *The Rise of the Network Society, Vol. 1*. Oxford: Wiley-Blackwell.

Crasborn, O., Sloetjes, H. (2010). Using ELAN for annotating sign language corpora in a team setting. In *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, LREC 2010.*

De Vos, C., Pfau, R. (2015). Sign language typology: The contribution of rural sign languages. In *Annual Review of Linguistics*, 1, 265-288.

Fenlon, J., Schembri, A., Johnston, T. & Cormier, K. (2015). Documentary and corpus approaches to sign language research. In Orfanidou, E., Woll, B. & Morgan, G. (Eds.), *The Blackwell guide to research methods in sign language studies.* Oxford: Wiley-Blackwell, 156-172.

Frishberg, N., Hoiting, N. & Slobin, D.I. (2012). Transcription. In Pfau, R., Steinbach, M. & Woll, B. (Eds.), *Sign language. An international handbook*. Berlin: De Gruyter Mouton, 1045-1075.

Geraci, C., Battaglia, K., Cardinaletti, A., Cecchetto, C., Donati, C., Giudice, S. & Mereghetti, E. (2011). The LIS Corpus project: A discussion of sociolinguistic variation in the lexicon. In *Sign Language Studies*, 11(4), 528-574.

Gertz, G., Boudreault, P. (Eds.) (2016). *The SAGE Deaf studies encyclopedia*. London: SAGE Publishing.

Goodstein, H. (Ed.) (2006). *The Deaf Way II Reader*. Washington, DC: Gallaudet University Press.

Hauser, P.C., Paludneviciene, R., Riddle, W., Kurz, K.B., Emmorey, K. & Contreras, J. (2016). American Sign Language comprehension test: A tool for sign language researchers. In *The Journal of Deaf Studies and Deaf Education*, 21(1), 64-69.

Humphries, T., Kushalnagar, P., Mathur, G., Napoli, D.J., Rathmann, C. & Smith, S. (2019). Support for parents of deaf children: Common questions and informed, evidence-based answers. In *International Journal of Pediatric Otorhinolaryngology*, 118, 134-142.

Legg, J. (2016). Exploring the promise of digital deaf histories. In *Sign Language Studies*, 17(1), 42-58.

Leigh, I.W., Andrews, J.F. & Harris, R.L. (Eds.) (2018). *Deaf culture*. San Diego, CA: Plural Publishing.

Lucas, C., Bayley, R. & Valli, C. (2001). *Sociolinguistic variation in American Sign Language*. Washington, DC: Gallaudet University Press.

Marschark, M., Tang, G. & Knoors, H. (Eds.) (2014). *Bilingualism and bilingual deaf education*. Oxford: Oxford University Press.

Mann, W., Haug, T. (2014). Mapping out guidelines for the development and use of sign language assessments: Some critical issues, comments and suggestions. In Quinto-Pozos, D. (Ed.), *Multilingual aspects of signed language communication and disorder*. Bristol: Multilingual Matters, 123-139.

Mayberry, R.I., Kluender, R. (2018). Rethinking the critical period for language: New insights into an old question from American Sign Language. In *Bilingualism: Language and Cognition*, 21(5), 886-905.

McCaskill, C., Lucas, C., Bayley, R. & Hill, J. (2011). *The hidden treasure of black ASL – its history and structure*. Washington, DC: Gallaudet University Press.

Monaghan, L., Schmaling, C., Nakamura, K. & Turner, G.H. (Eds.) (2003). *Many ways to be deaf: International variation in deaf communities*. Washington, DC: Gallaudet University Press.

Mougeon, R., Nadasdi, T. (1998). Sociolinguistic discontinuity in minority language communities. In *Language*, 74(1), 40-55.

Nyst, V. (2012). Shared sign languages. In Pfau, R., Steinbach, M. & Woll, B. (Eds.), *Sign language: An international handbook*. Berlin: De Gruyter Mouton, 552-574.

Padden, C.A., Humphries, T. (2005). *Inside deaf culture*. Cambridge, MA: Harvard University Press.

Parasnis, I. (Ed.) (1996). *Cultural and language diversity and the Deaf experience*. Cambridge: Cambridge University Press.

Perniss, P., Pfau, R., Steinbach, M. (2007). Can't you see the difference? Sources of variation in sign language structure. In Perniss, P., Pfau, R. & Steinbach, M. (Eds.), *Visible variation: Comparative studies on sign language structure*. Berlin: Mouton de Gruyter, 1-34.

Quer, J., Cecchetto, C., Donati, C., Geraci, C., Kelepir, M., Pfau, R. & Steinbach, M. (Eds.) (2017). *SignGram blueprint: A guide to sign language grammar writing*. Berlin: De Gruyter Mouton.

Quer, J., Steinbach, M. (2019). Handling sign language data: The impact of modality. In *Frontiers in Psychology,* 10:483. *doi: 10.3389/fpsyg.2019.00483.*

Rayman, J. (1999). Storytelling in the visual mode: A comparison of ASL and English. In Winston, E. (Ed.), *Storytelling & conversation: discourse in deaf communities*. Washington, DC: Gallaudet University Press, 59-82.

Rutherford, S. (1993). *A study of American Deaf folklore*. Burtonsville, MD: Linstok Press.

Sandler, W., Lillo-Martin, D. (2006). *Sign languages and linguistic universals*. Cambridge: Cambridge University Press.

Spencer, P.E., Marschark, M. (2003). Cochlear implants: issues and implications. In Marschark, M., Spencer, P.E. (Eds.), *Oxford handbook of deaf studies, language, and education*. Oxford: Oxford University Press, 434-448.

Supalla, T. (2001). Making historical sign language materials accessible: A prototype database of early ASL. In *Sign Language & Linguistics*, 4, 285-297.

Supalla, T., Clark, P. (2015). *Sign language archaeology*. Washington, DC: Gallaudet University Press.

Sutton-Spence, R. (2010). The role of sign language narratives in developing identity for deaf children. In *Journal of Folklore Research*, 47(3), 265-305.

Woll, B., Ladd, P. (2003). Deaf communities. In Marschark, M., Spencer, P.E. (Eds.), *Oxford handbook of deaf studies, language, and education*. Oxford: Oxford University Press, 151-163.

Zeshan, U., Palfreyman, N. (2017). Sign language typology. In Aikhenvald, A.Y., Dixon, R.M.W. (Eds.), *The Cambridge handbook of linguistic typology*. Cambridge: Cambridge University Press, 178-216.

Acknowledgments

This contribution has been made possible thanks to the SIGN-HUB project, which has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 693349. Josep Quer's contribution is also made possible by the Spanish Ministry of Economy, Industry and Competitiveness and FEDER Funds (FFI2015-68594-P), and to the Government of the Generalitat de Catalunya (2017 SGR 1478).