SIGN HUB

European Commission | Horizon 2020 European Union funding for Research & Innovation

Project Number: 693349

Annotation in ELAN

Roland Pfau

(with Carlo Cecchetto, Jens-Michael Cramer, Aslı Göksel, Jana Hosemann, Jordina Sánchez Amat, Rose Stamp, Süleyman Tasçi)

Version 1.1 – 31/07/2018

Project title: The SIGN-HUB: preserving, researching and fostering the linguistic, historical and cultural heritage of European Deaf signing communities with an integral resource

Lead contractor: Universitat Pompeu Fabra

Contact person:

Josep Quer
Departament de Traducció i Ciències del Llenguatge
Roc Boronat, 138
08018 Barcelona
Spain

Tel.   +34-93-542-11-36
Fax.   +34-93-542-16-17

E-mail: josep.quer@upf.edu

Work package: WP 2

Affected tasks: Task 2.4

Nature of deliverable[1]	R	DEM	DEC	O
Dissemination level[2]	PU	PP	RE	CO

COPYRIGHT

© COPYRIGHT SIGN-HUB Consortium consisting of:

- UNIVERSITAT POMPEU FABRA Spain

- UNIVERSITA' DEGLI STUDI DI MILANO-BICOCCA Italy

- UNIVERSITEIT VAN AMSTERDAM Netherlands

- BOĞAZICI ÜNIVERSITESI Turkey

- CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE France

- UNIVERSITÉ PARIS DIDEROT - PARIS 7 France

- TEL AVIV UNIVERSITY Israel

- GEORG -AUGUST-UNIVERSITÄT GÖTTINGEN Germany

- UNIVERSITA CA' FOSCARI VENEZIA Italy

- CONSORZIO INTERUNIVERSITARIO NAZIONALE PER L'INFORMATICA Italy

Version 1.1 – 31/07/2018

History of changes

VERSION	DATE	CHANGE	AUTHOR(S)/ REVIEWER(S)
1.0	23-07-2018	Initial version	Roland Pfau
1.1	30-07-2018	Revised version	A: Roland Pfau; R: Carlo Cecchetto, Carlo Geraci, Jana Hosemann, Jordina Sánchez Amat, Rose Stamp

CONTENTS

1. Scope of the document 5

2. Introduction 6

3. Overview 7

   3.1. Procedure 7

   3.2. Quantitative overview 8

   3.3. Challenges 8

4. Annotation per country 9

   4.1. France 9

   4.2. Germany 9

   4.3. Israel 10

   4.4. Italy 10

   4.5. Netherlands 10

   4.6. Spain 11

   4.7. Turkey 11

5. Outlook 13

1. Scope of the document

In this document, we provide an overview of the fragments of interviews (Germany, Italy, The Netherlands, Spain, Turkey) and narratives (Israel) that have been annotated in ELAN.

Following a brief introduction of the goal of this endeavor in Section 2, we sketch the annotation procedure and provide a quantitative overview in Section 3. In Section 4, we provide more detailed information concerning the annotated materials per country in the form of tables.

2. Introduction

ELAN (EUDICO Linguistic Annotator; https://tla.mpi.nl/tools/tla-tools/elan/) is a freely available annotation tool that is commonly used in sign language and gesture research, as it allows for the time-aligned annotation of videos. Both the video and the annotation are simultaneously visible on the screen. Furthermore, depending on the research question, ELAN allows for the individual creation of annotation tiers. In sign language research, for instance, it is common to create a tier for manually articulated signs, another one for non-manual signals, and a translation tier. Tiers are searchable, which makes ELAN a convenient tool for linguistic research, be it of a lexical or grammatical nature.

In the following, when we speak of "annotation", we mean annotation **beyond** the translation/subtitle tiers that have been created in the context of the subtitle task (Deliverable D2.25), that is, tiers on which aspects of the linguistic production of signers (interviewees) are glossed (i.e. tiers for manual signs, non-manual productions, and possibly comments).

The ultimate goal of the procedure is to perform linguistic analysis on the annotated materials, for instance, in order to detect instances of diachronic change (e.g. lexical or word order change). To that end, comparisons to the productions of younger signers will be informative (e.g. if corpus data are available for the respective sign language). Comparisons can be made within and across countries. Note that a linguistic analysis targeting a specific research question is not linked to a deliverable. Undoubtedly, however, such analysis carries enormous potential and is thus certainly part of our long-term goals.

3. Overview

3.1. Procedure

For the most part, annotation has been performed by Deaf researchers; in some countries, hearing researchers, who are fluent signers, also participated in the annotation task. In Germany, for instance, the first round of annotation is done by a hearing researcher, and this pre- annotation is then checked by a Deaf researcher. In general, for the annotation, one or more of the previously subtitled fragments have been selected, but there are some exceptions, where annotation has been performed on a fragment that had not previously been subtitled.

Annotation has been done in ELAN. The procedure is as follows: The video fragment is opened in ELAN, and tiers for annotation are created. It has been decided to minimally have – in addition to the translation tier – separate tiers for the right and the left hand, with glosses in the local spoken language, as these will be the most important tiers for linguistic analysis (e.g. lexical and phonological changes). Of course, optionally, additional tiers can be added for further investigation (see Section 4). Note that in some cases, i.e. sign languages for which a corpus is available, pre-defined ID-glosses for signs may exist. Given that double-checking ID glosses is an extremely time-consuming task, it was left open whether these ID-glosses should be used in the annotation.

Below we provide an example from a German Sign Language signer. The topmost tier containing information is the translation (German) tier (created in the context of the subtitling task); the next two tiers are the gloss tiers for the left hand (LH) and the right hand (RH), respectively. For this particular annotation, Germany created additional tiers, including a tier for non-manual markers (NMM) and tiers related to the function of a specific sign.
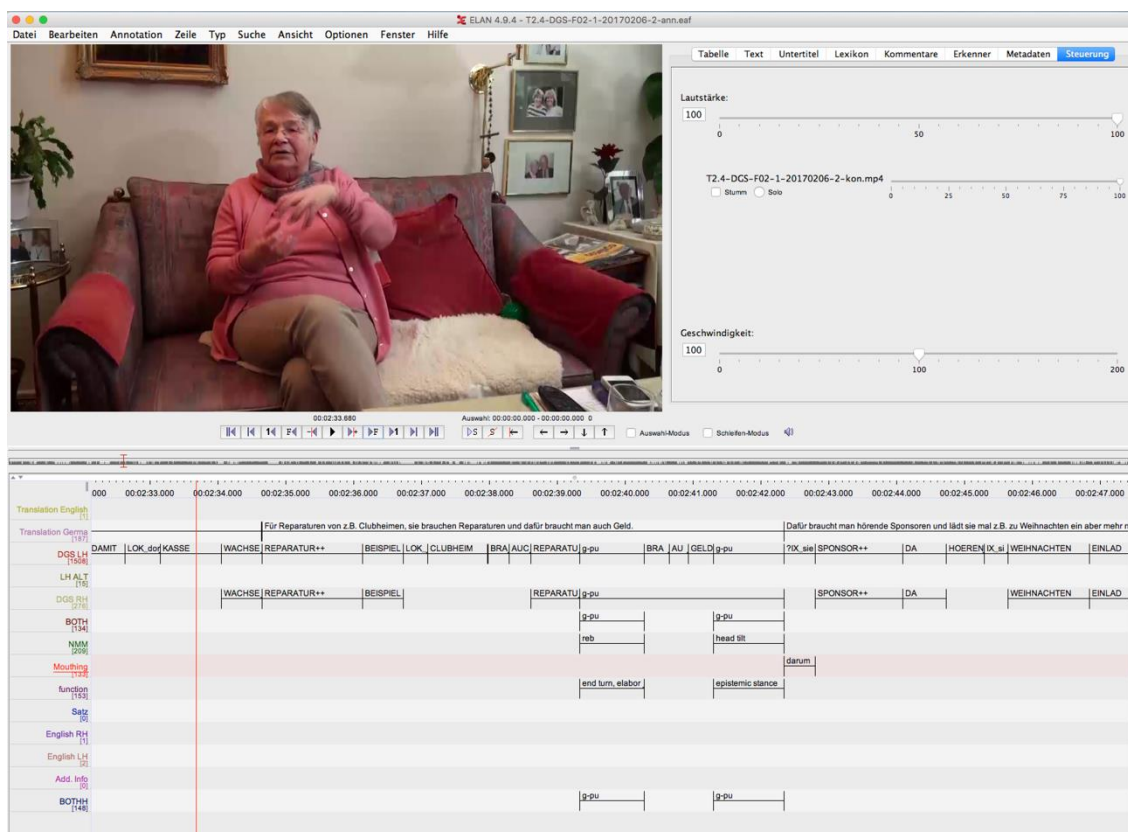
Given the stage of the annotation process (for some countries they are still under revision), it was decided not to include them as an appendix in this document; rather, the annotations will be made available via a Google Drive folder:

https://drive.google.com/drive/folders/1wCS_IsvXNiuN_h-BpCBpO9UdQwjyzlet?usp=sharing

3.2. Quantitative overview

In total, more than five hours of footage have been annotated. As is evident from the overview below, as well as the tables in Section 4, the amount of annotated materials differs between countries. As with subtitling (Deliverable D2.25), these quantitative differences result from differences with respect to budgeting of the task per country, and different manpower due to different costs for research assistants.

Remember that, in the context of this deliverable, annotation refers to the annotation of linguistic productions of signers, not to translations. In France and in Israel, no interviews have been conducted, and subtitling concerns previously existing materials: documentary films (France) and narratives (Israel). At present, no annotation activities are planned for these two countries, beyond the annotation that has been (and will be) done in the context of the subtitling task (translation tier; see Deliverable D2.25, Sections 4.1 and 4.3). In both countries, the budget that has been obtained will be devoted to subtitling. Still, we include both countries in the overview in Section 4.

The amount of footage annotated per country is as follows (hrs:min:sec):

- Germany: 03:18:31
- Italy: 00:30:00
- The Netherlands: 00:05:49
- Spain: 00:21:55
- Turkey: 00:55:41
  Total   **05:11:56**

3.3. Challenges

Annotation in ELAN is a time-consuming task, but the challenges that were encountered did not go beyond the challenges that generally characterize annotation work. These include (i) defining the onset and offset of a sign; (ii) glossing lexical variants (often done by specifying glosses, such as SIGN-A vs. SIGN-B); (iii) glossing of pointing signs (usually glossed as INDEX with extension for locus specification).

4. Annotation per country

As mentioned previously, different amounts of footage have been annotated per country for this deliverable. The following subsections are organized by country (listed in alphabetical order). In each subsection, the table provides an overview of the materials that have been annotated to date.

In Section 3.1, we pointed out that, besides the subtitling/translation tier created in the context of the subtitling task, annotation involved at least separate tiers for the right hand and left hand of the interviewee. Additionally, in some of the countries local glossing conventions were followed. In Spain, for instance, for LSC, the annotation conventions used for the LSC Corpus were adopted. Annotations in the local language will be used by the local community, and thus the obvious advantage of this strategy is that it ensures comparability with the local corpus data.

4.1. France

In France, the situation is different from that in the other participating countries. Rather than interviews with or narratives from Deaf elderly signers, in France, existing documentary movies will be subtitled. Consequently, subtitling does not only concern signing, but also spoken French. The main goal of annotation in ELAN is to extract French subtitles; that is, there will be no annotation beyond the tier to be annotated in the context of the subtitling task (see Deliverable D2.25 for details).

4.2. Germany

In Germany, fragments of six interviews (distributed over nine videos) have been annotated. In addition to the required tiers (see Section 3.1), a tier for the right hand with glosses in English has been created for some of the annotated fragments.

GERMANY		
INTERVIEW	DURATION	ANNOTATOR [α]
T2.4-DGS-F01-1-20170206-1	0:00 – 0:13	RA (HH)
T2.4-DGS-F01-1-20170206-2	0:00 – 26:02	RA (HH)
T2.4-DGS-F02-1-20170206-1	0:00 – 25:27	KH (HH)
T2.4-DGS-F02-1-20170206-2	0:00 – 25:08	KH (HH)
T2.4-DGS-F02-1-20170206-3	0:00 – 10:11	KH (HH)
T2.4-DGS-M05-1-20170512	0:00 – 28:32	FF (Gö), RA+KH (HH)
T2.4-DGS-F06-1-20170518	0:00 – 24:11	RA+KH (HH)
T2.4-DGS-M10-1-20170703	0:00 – 32:20	FF+RG (Gö)
T2.4-DGS-F21-1-20170928-2	0:00 – 26:27	RG (Gö)

[α] FF = Femke Feuker (Göttingen); RG = Rieke Giese (Göttingen); RA = Robin Angelini (Hamburg); KH = Katja Hopfenzitz (Hamburg)

Material annotated in ELAN – total time: 03:18:31 (hrs:min:sec)

4.3. Israel

In Israel, focus of annotation will be the subtitling (in English and Hebrew) of narratives produced by Deaf signers in Israeli Sign Language. That is, similar to France, there will be no annotation of the linguistic production of the signers (e.g. manual signs, non-manual markers). So far, subtitle annotations have been completed in Hebrew for a few narratives (see Deliverable D2.25).

4.4. Italy

In Italy, 6 interviews were chosen as a sample based on gender and geographical criteria: one male and female each from three different regions (northern, central, and southern Italy). For each interview, the same timeslot was extrapolated. In addition to the tiers common to all SIGN-HUB units, Italy added additional tiers concerning specific properties of Italian Sign Language, as produced by elderly signers, for instance, a tier for mouthings and tiers on which phonological and lexical changes can be specified. At present, gloss tiers (for both hands) have been annotated in Italian and English; also, the tiers for phonological/lexical changes and the translation have been filled.

ITALY		
INTERVIEW	DURATION	ANNOTATOR [a]
T2.4-MN1	10.00 – 15.00	RS
T2.4-GG1	10.00 – 15.00	RS
T2.4-FG1	10.00 – 15.00	RS
T2.4-SB1	10.00 – 15.00	RS
T2.4-TA1	10.00 – 15.00	RS
T2.4-ZR1	10.00 – 15-00	RS

[a] RS= Rita Sala

Material annotated in ELAN – total time: 00:30:00 (hrs:min:sec)

4.5. Netherlands

Given that the budget available for subtitling and annotation (which is external to the project budget anyway) was rather limited, it was decided to spend most of it on the subtitling task, as this is the task which is more crucial to the future goals of the project (documentary movie, platform). Therefore, only one fragment was annotated in ELAN. For this fragment, sign glosses were annotated in Dutch and English.

NETHERLANDS		
INTERVIEW	DURATION	ANNOTATOR [a]
T2.4-NGT-M04-1-20170222-1a	05:58 – 11:47	MvZ

[a] MZ = Merel van Zuilen

Material annotated in ELAN – total time: 00:05:49 (hrs:min:sec)

4.6. Spain

In Spain, three fragments from interviews in Catalan Sign Language have been annotated to date. In addition to tiers for the glosses for the right hand and the left hand, a comment tier has been created. Annotations have been created by Deaf students of the UPF LSC-teacher program. These training practices are being supervised by Delfina Aliaga.

SPAIN		
INTERVIEW	DURATION	ANNOTATOR [a]
T2.4-LSC-F05-1-20170921	3:32	DF
T2.4-LSC-M17-1-20171005	5:17	BF
T2.4-LSC-M13-1-20170922	13:06	EM & PF & RM

[a] BF = Berta Frigola; DF = David Falguera; EM = Edgar Murillo; PF = Pedro Frigola; RM = Ramon Munt

Material annotated in ELAN – total time: 00:21:55 (hrs:min:sec)

4.7. Turkey

In Turkey, one long fragment as well as 11 short fragments have been annotated in ELAN. Besides tiers for the right hand and the left hand (with glosses in Turkish), a combined glossed tier has been created.

TURKEY		
INTERVIEW	DURATION	ANNOTATOR [a]
T2.4-TİD-M06-k1	40:00	ETÖ, GY, YE
T2.4-TİD-GA01_Kültürel Miras - Aile	1:58	ETÖ
T2.4-TİD- GA01_Kültürel Miras - Eğitim k2	1:10	ETÖ
T2.4-TİD- GA01_Kültürel Miras - İletişim 2	0:33	ETÖ
T2.4-TİD- GA01_Kültürel Miras - İletişim 3	0:18	ETÖ
T2.4-TİD- GA01_Kültürel Miras - İletişim	0:32	ETÖ
T2.4-TİD- GA01_Kültürel Miras - İş Hayatı k2	0:34	ETÖ

TURKEY		
INTERVIEW	DURATION	ANNOTATOR α
T2.4-TİD- İA01_Kültürel Miras - Askerlik	3:30	ETÖ
T2.4-TİD- İA01_Kültürel Miras - Sinema	1:55	ETÖ
T2.4-TİD- SK01_Kültürel Miras - İletişimk1	1:22	YE
T2.4-TİD- SK01_Kültürel Miras - Eğitim-Okul 2	2:09	YE
T2.4-TİD- SK01_Kültürel Miras - Eğitim-Okul 3	1:40	YE

α ETÖ = Elvan Tamyürek-Özparlak; GY = Gül Yiğit; YE = Yusuf Ermez

Material annotated in ELAN – total time: 00:55:41 (hrs:min:sec)

5. Outlook

Depending on the country-specific budgets, and success in recruiting additional funding, the annotation work will continue.

- In Germany, external funding by the state of Lower Saxony allows the team to continue annotating sign glosses until the end of the project. Annotation will be done in a first step by hearing DGS-proficient student assistants (pre-annotation), followed by proofs of two deaf native DGS signers (final annotation).
- In Italy, annotation will continue with the goal of annotating about 5 minutes from each interview. The annotation will be guided by some research questions for which specific ELAN tiers have already been created. These include lexical and phonological changes and some syntactic properties like order between modals, aspectual markers and negation with respect to the main predicate.
- In the Netherlands, annotation work has been financed by project-external sources. These sources are now exhausted, and additional annotation will thus only be possible if external funding can be acquired.
- In Spain, annotation of excerpts of Catalan Sign Language interviews will continue during the next months, until one hour of video recording is annotated. Annotation of excerpts of interviews in Spanish Sign Language will begin in October 2018.
- In Turkey, subtitling is financed from the budget allotted to the Turkish SIGN-HUB team. As a next step, annotation will be continued with interviews with participants older than 80 years of age, as more age-related variation is expected in this age group. Furthermore, it is planned to translate the existing Turkish glosses into English; for the fragments that will be annotated from scratch, annotation will be done in Turkish and English.

As mentioned above, in the future, annotations will be crucial for conducting linguistic studies based on the interviews and narratives, in particular, studies that address issues of diachronic change (e.g. phonological change, lexical innovations). This would involve comparison of (selected aspects) of the productions of elderly signers to productions of younger signers (possibly based on available corpus data). Such studies could also be of a cross-linguistic nature, that is, they could include a comparison of diachronic changes as observed for different sign languages.

We also anticipate that issues of diachronic change will be addressed in one or more contributions to the edited volume (Deliverable 2.29) – be it as the main focus of a contribution or as an issue that comes up in the context of another topic, e.g. Deaf education.