DGS Corpus & Dicta-Sign: The Hamburg Studio Setup

Thomas Hanke, Lutz König, Sven Wagner, Silke Matthes
Institute of German Sign Language and Communication of the Deaf, University of Hamburg
{thomas.hanke,lutz.koenig,sven.wagner,silke.matthes}@sign-lang.uni-hamburg.de

Abstract

We describe the setup for a mobile studio used for data collection of both the DGS Corpus project and the DGS part of the Dicta-Sign corpus. This includes camera positioning and the software used to conduct a recording session as well as the rationale behind the decisions taken in that respect, but we also investigate the challenges the moderator of a session has to face.

1. Introduction

Not taking into account budget restrictions, the setup of a sign language studio always is a balancing act between high quality recordings on the one hand not to make the transcription process even more complicated than it is anyway and possibly to enable automatic processing of the recordings, and on the other hand an environment where the informants still feel comfortable enough so that the recording situation does not have too much impact on the signing. In the case of the DGS Corpus project, an additional constraint is that the studio is to be relocated twelve times over the course of two years as it was decided to make the recordings in the regions instead of inviting participants to one central place to avoid dialectal mixing. One of the implications of this approach is that the studio is operated by non-specialist deaf fieldworkers with limited time available for training.

2. Overall Setup

The elicitation setting for both DGS Corpus and Dicta-Sign involves two informants interacting in different ways with each other and a moderator (the fieldworker from the region) leading the session.

During a recording session the pair of informants is sitting facing each other at an approximate distance of three meters. Camera positions in the studio guarantee that each informant is being filmed separately while there exists another camera to film the overall scene. Elicitation material to be shown to the informants is presented on screens in the middle between the signers very close to the ground in order not to interfere with their views of each other. The moderator is sitting next to the informants as shown in the picture below. He/she introduces the tasks and observes the conversation, but only interferes with the conversation if absolutely necessary. Monitors in front of the moderator display the "Session Director" (described below) as well as what can currently be seen on the informants' screens.

3. Camera Positions

The camera setup we finally ended up with consists of seven cameras altogether, three on each informant and one for the whole scene including the moderator. Two HD cameras provide frontal views of the informants while birds-eye cameras capture each informant's signing from above to help human transcribers to interpret the signing. Additionally two stereo cameras mounted on top of the frontal-view cameras capture the signing in parallel. They are to provide footage that allow image analysis to reconstruct 3D information and help automatic processing. The seventh camera is again an HD camera which captures the whole scene, i.e. both informants as well as the moderator interacting with the informants to give the transcriber a quick overview and to help him/her to exactly identify interactions between the three participants.

One of the advantages of using HD cameras is that extra cameras for close-ups, e.g. on the signers' faces, are not necessary as the spatial resolution of the those parts of an HD image still is comparable to what can be achieved with SD cameras – without the need to track positions.

In earlier projects the informants were also seated opposite each other but at a slight angle. The cameras were positioned at eye height on the side of each informant. The drawback of this approach is some informants constantly target their signing back and forth between the addressee and the camera, which makes it difficult to identify important aspects as body shifts in the discourse situation. Pre-tests were therefore conducted testing different seating arrangements and camera positions. It was found that a setting is optimal where the informants are directly facing each other in order to assure a relaxed conversational situation. With this, three different camera positions are possible: on the side or above the other informant's head, or between the informants. The last option requires a large distance between the informants or shooting at wide angle. In

addition, to avoid blocking the informants' views, the cameras would need to be mounted quite low, having to shoot upwards, something our transcribers did not like at all. The side position provides a view on the informant at eye height but at a side angle, which makes body shift and eye gaze tracking difficult for the transcriber as well as for semi-automatic processing. For the Hamburg studio setup it was therefore decided to use a setting with the main cameras positioned above and behind the heads of the informants, which provides a view from the front but with a slight angle from above. The pre-tests revealed that with a distance of approximately three meters, the distortion introduced by the elevated position of the camera does not negatively affect the transcription from video. Instead, this setting provides a front view of the informant similar to the addressee's, allowing identification of body shifts as well as eye gaze direction more easily. At the same time, especially with the monitors being located on the floor between them, the informants did not feel this to be an unnatural distance for their interaction.

4. Procedure of Elicitation Sessions

4.1 The moderator's role and presentation of the stimuli

The fieldworker of a certain region also serves as the moderator during the elicitation sessions. In order to avoid influences on the language production it is crucial that the moderator is a Deaf person and that no other (hearing) person is present in the studio throughout the session.[1] The moderator is responsible for the presentation of the individual tasks as well as a smooth run of the whole session.

Each task is briefly introduced by the moderator, followed by a detailed explanation and instruction for the informants, presented in a DGS video clip on the informants' monitors. While these pre-recorded explanations ensure that all informants get exactly the same information and that nothing is left out, further clarifications given by the moderator might become necessary for some of the informants before starting a task. The materials used as stimuli during a task comprise of different media formats, including pictures, drawings and video clips. They are shown as slides on the informants' monitors in a semiautomatic presentation, partly one slide following the other at a fixed speed, partly controlled by the moderator (e.g. allowing for in-between questions by the informants). Depending on the individual task the presentation of this material might be identical or different for the two informants.

The aim of the tasks and the stimuli used is to evoke a conversation between the two informants, while the moderator should observe the conversation and only interfere if absolutely necessary. However, it is the moderator's duty to check the time used for each task in order to ensure that enough time is left for the rest of the sessions. For both Dicta-Sign and DGS corpus extra tasks were planned that can be included or left out depending on the time left.

Leading the elicitation sessions and taking care of every aspect required leaves a heavy responsibility with the moderator. Training sessions are therefore required to fulfil this task. However, with limited time available prior to the elicitation and especially with long elicitation sessions to be performed (seven hours for the DGS corpus elicitation), a (semi)automatic control of the session should be implemented wherever possible. A custom software was therefore developed in order to support the moderator in his/her work and to ensure a smooth work flow.

4. 2 Session Director

Session Director allows the moderator to present slides to the informants. The screen configuration for the presentation of slides is one screen facing each informant and for the moderator one screen to run Session Director and two additional screens to observe the presentations on the informants screens. This means that the moderator can also see what is shown to the informants.[2]

In addition to the list of tasks, Session Director's main window shows detail of the task current worked on. This includes a progress bar showing the time already spent on the task in relation to the time planned in as well as the sequence of subtasks, such as the instructions to be presented to the informants, stimuli presentations and conversations between the informants. While the sequence of events is predefined, it is the moderator clicking start buttons on the screen to activate the next step. This allows the moderator to check if all explanations have been understood. In case, s/he can decide to repeat instructions on screen or to rephrase the task himself/herself before moving on.

All the moderator's interactions with Session Director are logged with time stamps. This allows us to determine automatically where on the videos certain tasks (or pauses) can be found and also to conclude from our knowledge of the tasks who presumably is the active signer at a given point in time and e.g. to use this information to automatically zoom the displayed video onto that person.[3,4]

It is neither doable nor desirable to keep each task to the planned duration. It is, however, necessary to keep the total session time close to the plan as the informants may need to catch their train back home at the end of the day etc. For this purpose, Session Director shows another window giving the actual time, the elapsed time in the session and the time before/behind plan. The window changes colours to signal when deviations exceed a

[1] A Deaf technician monitors the recording equipment from next door. While the informants know that, they do not see him/her during the session.

[2] Technically, the informants' screens are computers (iMac 24") with a second monitor mirroring the content to the moderator. They run a slide show presentation program (Apple Keynote) remote-controlled (via AppleEvents) by Session Director running on the moderator's iMac. The integration of Keynote allows us to use its full repertoire of features such as transition effects where appropriate.

[3] Notes that the moderator takes during the session (lower right of the main window) are also output into that log file. While we did not expect this to be much used, one of the three moderators did use this possibility regularly.

[4] The log also easily translates into tagging in our transcription environment iLex, allowing links from the transcript to the task description and vice versa.

certain threshold and require corrective action. If the session is well behind for some reason, it might no longer be possible or advisable to reduce the signing time for each task to a minimum, but to skip entire subtasks or even whole tasks. Session Director supports the moderator in these decisions with marking lower-ranked tasks that could be skipped.
In general, the order of tasks, including breaks, is pre-determined by the session description. The moderator has, however, the freedom to rearrange tasks or to change the expected duration of the session Session Director measures progress against. This might for example become necessary if one informant arrives late. But we have also experienced the opposite: Informants had so much fun with the tasks that the session was well behind, but they preferred to stay an hour longer in order not to miss any of the tasks still to come. With the moderator redefining the session duration, Session Director went back from "condition red" to "condition green" so that the moderator could relaxedly monitor progress without being constantly reminded to rush or skip optional parts.

4. 3 Session Description Files

When launched, Session Director loads an XML file describing the session. For each task and subtask, it defines the expected and maximum acceptable duration, the text of the user interface elements visible in Session Director and of course the ids of the slides to be shown to the informants, either as a common set or separately for each informant. Furthermore, it defines the relative importance of each task which Session Director will eventually use to mark tasks that can be skipped. In addition, text can be entered that will be displayed alongside with the task detail. This could be reminders to the moderator what questions to ask to get a discussion going should that turn out to be necessary for a specific task.
Separating the description from the program also makes it easy for us to produce session-specific files. We currently use this to arrange for alternating tasks: Some tasks are only used in every second or fourth session.[5] The moderator can then use the file produced for a specific session to prepare for the session, e.g. by adding keywords to indicate the informants' hobbies and such, information to be used in the warm-up phase or whenever some intervention by the moderator becomes necessary.

4. 4 Training

While the Session Director user interface is quite straightforward to use, time management for the data collection sessions remains a demanding task and requires that the moderator is familiar with the program under all conditions. Moderators are introduced to the program within the fieldworkers training courses. Right away, they have to use the program to manage rehearsal sessions with a mixture of cooperative and not-so-easy "informants". Time management remains the most complex aspects, and feedback from the moderators has led to some modifications as to when the colour-coded time reminders appear. Feedback from the first sessions also made us introduce a Pause function to halt the task time should a spontaneous break become necessary (while of course the session clock is not stopped).
The fieldworker's manual also documents Session Director as well as the essentials of time management. In addition, the slides as well as training version of Session Director[6] are made available to the moderators.

5. Technical Details

The HD cameras we use are 3 Sony EX3 and 2 Sony HRX-MC1P (for the birds-eye views), all recording at 1080i25. These cameras store data locally on memory cards (only used as a fallback solution) and at the same time stream into MacPros running FinalCutPro (via SDI connections for the EX3 and HDMI connections for the HRX-MC1P.)
The stereo cameras mounted on top of the EX3s are PointGrey Bumblebee 2 models capturing 640x480p48 for each channel. They are connected via FireWire to MacPros running capture software provided by Dicta-Sign partner University of Surrey (currently under Windows XP).[7]
With total data rates of 700 GBytes per hour, we are not able to transfer the data to the Hamburg server before the next session starts. Instead, we swap hard disks (Raids consisting of two 2TBytes hard disks for each computer) after each fourth session and transport them back and forth in special suitcases. (Local backup is available just in case.)
For the mobile studio in the described configuration, we need a room of at least 5m x 5m optimally with a ceiling height of at least 3m and a smaller room next door. However, the larger the room, the less packed it is, the easier it is to make the informants feel comfortable.
For relocating the studio, we have transport cases for all equipment including spare parts (including one HRX-MC1P, one MacPro, one iMac). Deinstallation, actual transportation and installation at the next site is organised by an external service provider.

Session Director is available free of charge for MacOS X only as it heavily relies on MacOS-specific functionality.[8] Source code is available upon request. As most of the user interface of Session Director is determined by the session description XML files, no localisation is necessary. The manual currently is available only in German, but an English version should become available in the future.

[5] The Keynote slides document is the same for all variations of the session: It contains the material for all variations, with some not being called in each session.

[6] The training version does not actually remote-control the slides computers as that would require a multi-computer setup at the moderator's home.

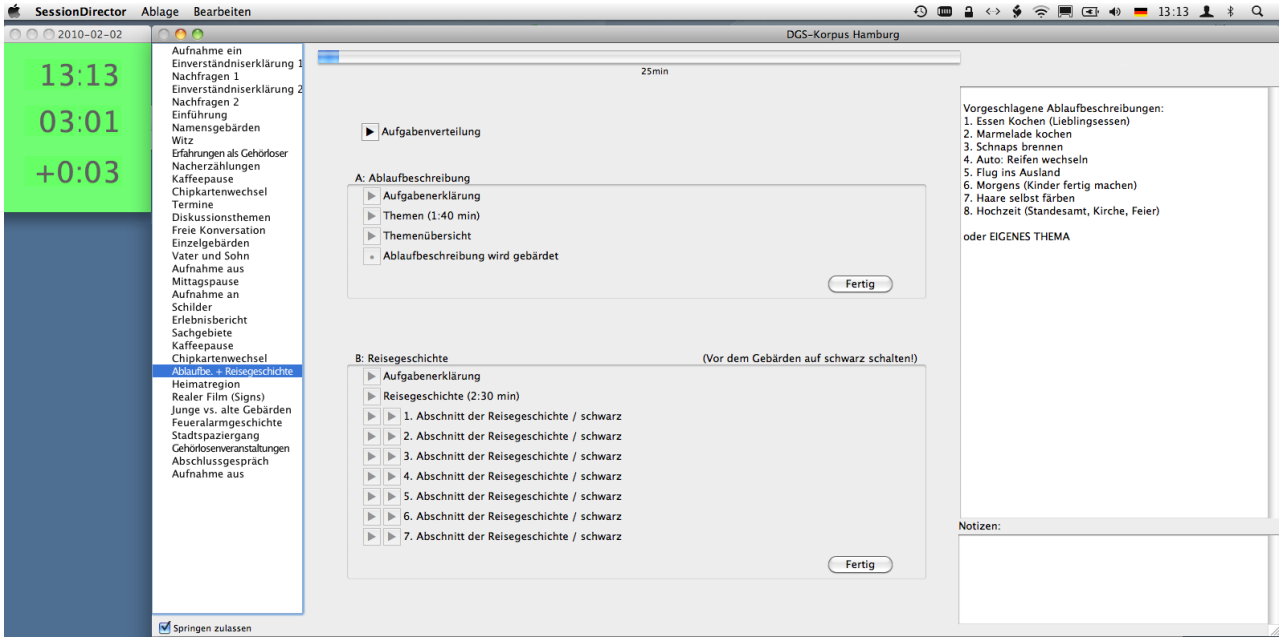
[7] Remote control for FinalCutPro and the Bumblebee recording software was implemented and integrated into Session Director but is currently not used. It turned out that too many things can happen that need immediate attention than could be handled by the moderator. We therefore decided to have a technician on-site which allows the moderator to concentrate on dealing with the informants and managing the session.

[8] The training version, however, is also available for Windows.

Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 231135 and from the German Academies of Science Programme.

Example: Excerpt from a Session Description XML file and the corresponding view in Session Director:




```
<session name="DGS-Korpus Hamburg" standard_rank="5" green_threshold="-30" yellow_threshold="0" red_threshold="20" maximum_duration="420"
    keynote_a="a.local" keynote_b="b.local" >

    <recording id="start 1" name="Aufnahme ein" action="start" average_duration="1">
        <checklist text="Alle Kameras nehmen auf"/>
        <checklist text="Wichtig: Über den Monitoren einmal laut Klatschen (zur Filmsynchronisation)!"/>
    </recording>

    ...

    <task id="27" name="Ablaufbe. + Reisegeschichte" average_duration="25" maximum_duration="30" rank="2" >
        <presentation id="27.1.1" name="Aufgabenverteilung" source_a="314" source_b="316" duration="0:33" />
        <subtask id="27.2" name="Ablaufbeschreibung" target="A" >
            <presentation id="27.2.1" name="Aufgabenerklärung" source_a="318" source_b="332" duration="0:41" />
            <presentation id="27.2.2" name="Themen (1:40 min)" source_a="320" source_b="430" duration="1:39" />
            <presentation id="27.2.3" name="Themenübersicht" source_a="322" source_b="433" average_duration="2" maximum_duration="2" />
            <narration id="27.2.4" name="Ablaufbeschreibung wird gebärdet" source="2" average_duration="7" maximum_duration="9" />
        </subtask>
        <subtask id="27.3" name="Reisegeschichte                                    (Vor dem Gebärden auf schwarz schalten!)" target="B" >
            <presentation id="27.3.1" name="Aufgabenerklärung" source_b="330" source_a="328" duration="0:33" />
            <presentation id="27.3.2" name="Reisegeschichte (2:30 min)" source_b="334" source_a="421" duration="2:26" />
            <narration_with_stimulus id="27.3.3" name="1. Abschnitt der Reisegeschichte / schwarz" source_b="353;2"
                source_a="2" average_duration="1:00" maximum_duration="1:17" />
            <narration_with_stimulus id="27.3.4" name="2. Abschnitt der Reisegeschichte / schwarz" source_b="356;2"
                source_a="2" average_duration="1:00" maximum_duration="1:17" />
            <narration_with_stimulus id="27.3.5" name="3. Abschnitt der Reisegeschichte / schwarz" source_b="359;2"
                source_a="2" average_duration="1:00" maximum_duration="1:17" />
            <narration_with_stimulus id="27.3.6" name="4. Abschnitt der Reisegeschichte / schwarz" source_b="362;2"
                source_a="2" average_duration="1:00" maximum_duration="1:17" />
            <narration_with_stimulus id="27.3.7" name="5. Abschnitt der Reisegeschichte / schwarz" source_b="365;2"
                source_a="2" average_duration="1:00" maximum_duration="1:17" />
            <narration_with_stimulus id="27.3.8" name="6. Abschnitt der Reisegeschichte / schwarz" source_b="368;2"
                source_a="2" average_duration="1:00" maximum_duration="1:17" />
            <narration_with_stimulus id="27.3.9" name="7. Abschnitt der Reisegeschichte / schwarz" source_b="371;2"
                source_a="2" average_duration="1:00" maximum_duration="1:17" />
        </subtask>
    </task>

    ...

    <comment id="27">
Vorgeschlagene Ablaufbeschreibungen:
1. Essen Kochen (Lieblingsessen)
2. Marmelade kochen
3. Schnaps brennen
4. Auto: Reifen wechseln
5. Flug ins Ausland
6. Morgens (Kinder fertig machen)
7. Haare selbst färben
8. Hochzeit (Standesamt, Kirche, Feier)

oder EIGENES THEMA
    </comment>

    ...

</session>
```