

Transcription guidelines for the Oral History Project

Scope

Transcription of (video) interviews conducted as part of the OHP

Goals

1. Verbatim and complete transcription of all interviews so that the transcripts can be used for content analysis by the authors of the various chapters of the OHP book;
2. Basic transcription of all verbal contents of the interviews so that the transcripts can be augmented later with additional information on pauses, nonverbal interaction, and context and with additional codes and code tiers for specific analytical purposes. None of these are part of the transcription at this stage.

Compatibility

At this very basic level, the transcripts need to be compatible with the guidelines from the *Gesprächsanalytisches Transkriptionssystem 2* (Selting et al. 1998, 2009) and the CHAT transcription format (MacWhinney, 2000).

Transcript computer files

The transcript can be created as a Word file (.doc, .docx, .rtf) or as a plain text file (.txt). The Word file must **not** contain any formatting, comments or tracked changes, images or drawings, or highlighting. Each interview's transcript needs to be in a single file.

Initial header sections

These 10 header lines are at the very beginning of your transcript. They start with @ and need to be entered in the order given below. Here they are explained; in the appendix is a template which you can simply copy into your transcription document and then edit it with the information from your interview.

@Begin

@Languages:

@Participants:

@ID:

@Media:

@Interaction Type:

@Location:

@Date:

@Time Duration:

@Transcriber:

@Begin

This is a bare line; there is no colon. It just signals the beginning of the transcript for the analysis software.

@Languages: deu, eng

Depending on which words were used in the interview, either one or both languages are listed with these acronyms. If words in other languages were used, write the language as a word using lower case only.

@Participants: PME Paul Meier Interviewee, MSC Mathias Schulze Interviewer

As in the example, all participants who spoke or are seen in the interview video are listed. List entries are separated by a comma. Each entry consists of three parts separated by a blank space only: a speaker ID of three capital letters (initial of the first name and the first two letters of the surname), the full name, and the role.

@ID: language|corpus|code|age|sex|group||role||

There must be one @ID line per interview participant, for example:

@ID: deu|ohp|PME|75|male|poland||Interviewee||

The number of bars | is important; we are leaving out some possible specification. Use the language acronym listed under @Languages, followed by 'ohp' for the corpus name, the speaker ID as above, the age in digits, 'male' or 'female', the

country or region of origin (e.g., poland, germany, austria, hungary, sudenten, transylvania, ...), and the role as given above.

@Media: meier, video

The first component is the filename of the video or audio file you are transcribing, without its file extension. The second part will normally be 'video' (small letters) to identify the file as a video file. If you are transcribing an audio-recorded [only](#) interview, state 'audio.'

@Interaction Type: interview

In our case, this will always say interview.

@Location: campus, Waterloo, ON, Canada

For the first component, either state 'campus' or 'home'. Then give the city, town, or village, followed by province and country.

@Date: 01-JUL-2013

Use this date format to record the date of the interview.

@Time Duration: 1:32:03

This gives the duration of the interview in hours, minutes, and seconds. If the interview was less than one hour state the duration like this: 0:52:00. Always use two digits for both minutes and seconds.

@Transcriber: Gerd Helbig, 16-MAY-2014

Give your name and the date of the last day you worked on the interview.

At the very end of the interview transcript is the line

@End

It works similarly to the @Begin line.

Transcribing

Main line

All uttered words are entered in the main line. Each main line starts with an asterisk followed by the speaker/participant ID, a colon, and a tab. Each main line ends with hard return (hit Return or Enter on your keyboard). There are no punctuation signs.

*PME: this is an example of a main line

Length of the main line

Each main line contains a maximum of one turn or one predication.

A **turn** is the interview segment spoken by one participant. A predication is a complete clausal unit. For example "lassen sie uns mit dem interview beginnen" as a simple sentence is a predication. So is the complex sentence "ich weiß dass sie 1961 nach kanada gekommen sind" because leaving out the subclause would render the main clause "ich weiß" incomplete and hence agrammatical. The complex sentence "ich erkläre das nur weil sie viel jünger sind als ich" consists of two predications: "ich erkläre das nur" and "weil sie viel jünger sind." Here the main clause is a complete syntactic unit because the subclause is syntactically optional.

- **Address phrases** are treated as predications, so "herr Meier" "können sie bitte etwas über ihre arbeit im deutschen klub sagen" are two predications and need to be on two consecutive main lines.

- **False starts** and anacolutha (unfinished clausal units) followed by a predication can be on the same main line.

- For **overlaps** (more than one participant speaking at the same time), each speaker's (partial) predication or turn is written on a separate main line.

Turns that consist of one predication only are on one main line; for turns with more than one predication, each predication is on a separate main line. The main line does not contain any punctuation and all words are spelled with small letters.

*MSC: lassen sie uns mit dem interview beginnen

- *PME: was für fragen haben sie denn
- *PME: ich kann ihnen erzählen wie ich nach Kitchener kam
- *PME: oder wollen sie das nicht wissen
- *PME: denn sie weil ich ihnen auch andere sachen erzählen kann
- *MSC: herr Meier
- *MSC: erzählen sie mir doch bitte wo sie herkommen und aufgewachsen sind

Transcribing words

- All words uttered during the recording are transcribed.
- All German words follow the German standard orthography, and all English words follow English orthography. However, all words that are not **proper nouns** (Eigennamen) have small letters only. Proper nouns start with a capital letter; compound proper nouns such as “Bundesrepublik_Deutschland” are linked with an underscore.
- Filled **pauses**, which are not dictionary words (e.g., ähm, mmhm) should be transcribed, if possible. Approximations of what is heard suffice. Silent pauses do not need to be marked.
- If inflections are clearly **elided**, as in “ich hab,” the words can be spelled like this. Do not use an apostrophe. Often speakers drop vowels from unstressed syllables. These can be added in parentheses, e.g. “spiel(e)n” “(a)bout.”
- Do use an apostrophe in English transcriptions if it is standard, e.g., it’s, i’m, Bob’s, they’re, etc.
- **Deviations from standard German/Canadian English** (e.g. inappropriate articles or inflections such as “mit die pferde” or “mit rote haare”) are to be transcribed as they are.
- If the speaker left out words (a missing article or preposition, for example: “ich war in Schweiz zwei Jahre”), they must not be added.
- All **numbers** are to be given as words not digits: “zweiundzwanzig prozent,” “zwei dollar fünfzig”—exactly as they were pronounced. Thus, 1965 can be spelled “neunzehnhundert fünfundsechzig,” “neunzehnhundertundfünfundsechzig,” “neunzehn fünfundsechzig” or even “fünfundsechzig,” depending on what the speaker said.
- **Acronyms and abbreviations** are spelled as such, e.g., “tv,” “dvd.” Proper noun acronyms (which are pronounced as a word) such as “Unesco” or “Tüv” are spelled with a capital initial and the rest small. Proper noun abbreviations (for which each letter is pronounced separately) are spelled as linked capitals, e.g., “U_S_A,” “B_R_D.”

No deidentification

All names of people, places, groups, institutions, clubs, etc. are to be given in the transcript, as they were said. We will need this concrete information for writing the various book chapters. The transcripts will be anonymized later and before they become part of the corpus for research purposes.

Unintelligible speech

Use the symbol xxx when you cannot hear or understand what the speaker is saying. If you believe you can distinguish the number of unintelligible words, you may use several xxx strings in a row, each xxx representing one word. Please place a time stamp after the xxx, e.g., 0:50:23.

- *PME: mein bruder war ein alter xxx und xxx xxx xxx

Dependent tiers

Underneath each main line can be further lines with additional information. These are called dependent tiers. Dependent tiers all start with a percentage sign %, followed by a three letter identifier, a colon and a tab.

Dependent tier: comments

Use this comment line to make comments when you are unsure about your transcription, provide a possible alternative, or simply indicate what you guessed. In these comments, you can also note that you observed relevant nonverbal behaviour, such as the speaker getting very sad or angry or his/her gaze or gestures augmenting or disambiguating the information in the text. It’s best to write the comments in plain English.

- *PME: mein bruder war ein alter xxx und xxx xxx xxx
- %com: Maybe first xxx is schlawiner; speaker pumping his fist in anger
- *PME: als kind wohnte ich in Buckelhof
- %com: I am not sure about this place name; speaker said this with great pride

Dependent tier: explanations

Later on authors of the various book chapters will search these transcripts for information on their topic. These topics are likely to emerge:

- childhood
- immigration
- acculturation (settling in)
- language
- education
- social life
- professional life
- social network
- war
- post-war
- ...

Time frames and periods are also going to be relevant:

- thirties
- forties
- fifties
- sixties
- ...
- 2000s

Countries and regions, too: siebenbürgen, österreich, ontario, quebec, ...

If the text in the main line clearly refers to one of the topics, time frames, or locales without mentioning them verbatim, this means we could not search for them. Then it would be good to list them in the explanation tier. If you notice anything interesting or out of the ordinary in the interview, write it down in the explanations tier as well, e.g., interesting=comment_on_kitchener. Use topic=, time=, or locale= with a suitable keyword after the equal sign. If you need more than one word as the keyword, link them with underscores as in the previous example.

*PME: als kind wohnte ich in Buckelhof

%exp: locale=vorpommern topic=childhood

PAYMENT

You should aim to produce a clean and complete transcript of 1 hour of video in 4 hours and 45 minutes. For the transcription of one hour of interview data, you will be paid \$130.00. Transcription of interview minutes over and above a full hour will be paid on a pro rata basis at about \$2.17.

How does it work? Request a video file from Lori (lstraus@uwaterloo.ca). Download it. Produce the complete transcript, proofread it thoroughly and send it to Lori. Please be sure to include your payroll or student number with your first transcription. She will arrange payment based on the length of the video file.

References

MacWhinney, B. (2000). The CHILDES Project: Tools for Analyzing Talk. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates. (also available at <http://childes.psy.cmu.edu/manuals/CHAT.pdf>)

Selting, M, P. Auer, B. Barden, J. Bergmann, E. Couper-Kuhlen, S. Günthner, ... S. Uhmman (1998). Gesprächsanalytisches Transkriptionssystem (GAT). Linguistische Berichte 173, 91-122. (also available at <http://www.medien-sprache.net/de/medienanalyse/transcription/gat/gat.pdf>)

Selting, Margret et al. (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion 10, 353-402. (available online at <http://www.gespraechsforschung-ozs.de/heft2009/px-gat2.pdf>)

Transcript template

@Begin

@Languages: deu, eng

@Participants: PME Paul Meier Interviewee, MSC Mathias Schulze Interviewer

@ID: deu|ohp|PME|75|male|poland||Interviewee||

@Media: meier, video

@Interaction Type: interview

@Location: campus, Waterloo, ON, Canada

@Date: 01-JUL-2013

@Time Duration: 1:32:03

@Transcriber: Gerd Helbig, 16-MAY-2014

*MSC: lassen sie uns mit dem interview beginnen

*PME: was für fragen haben sie denn

*PME: ich kann ihnen erzählen wie ich nach Kitchener kam

*PME: oder wollen sie das nicht wissen

*PME: denn sie weil ich ihnen auch andere sachen erzählen kann

*MSC: herr Meier

*MSC: erzählen sie mir doch bitte wo sie herkommen und aufgewachsen sind

*PME: mein bruder war ein alter xxx und xxx xxx xxx

%com: Maybe first xxx is schlawiner

*PME: als kind wohnte ich in Buckelhof

%com: I am not sure about this place name

%exp: locale=vorpommern topic=childhood

*PME: und da gings uns ganz gut

*MSC: wo ist denn Buckelhof

@End