Code Quality – Transcription and Analysis Guide

ITiCSE 2017 working group WG5

April 27, 2017

1 Main Contacts

- Daniel Toll, +46 ..., daniel.toll@lnu.se, Skype: ...
- Jürgen Börstler, +46 ..., jubo@acm.org, Skype: ...

2 Analysis Protocol

The WG-leaders will provide a shared spreadsheet for collecting all answers from Q1–Q3 and Q5–Q11. Initially, we will do descriptive statistics for these data and an occasional statistical test, in case we find something interesting and/or have enough datapoints.

Question Q4 will be transcribed and (see next section) and coded/analyzed at ITiCSE.

3 Transcription

Transcription is the act of translation from oral to written language and discourse. To get useful transcriptions, we need to make sure that everyone transcribes using the same process. This is especially important when using semi-automatic analysis tools (which we might try). This section describes the process of transcribing. This is also important since we might need to state explicitly in our report how the transcriptions were done.

Since we are not transcribing for conversational analysis but to get knowledge, we suggest to transcribe into fluent and readable style that can be sent back to the interviewees for confirmation (to increase reliability). "If the transcripts are to be sent back to the interviewees, rendering them in a more fluent style should be considered from the start" (see Brinkman and Kvale (2015)[1]).

To make the transcription more readable and fluent we remove words that have little or no meaning. For example we do not need to transcribe words like "eh" or words or phrases that are duplicated without particular meaning. Example: "I think that [eh] think that [eh] ..." would be transcribed into "I think that ...".

All transcriptions should be in English. If the actual interview was carried out in a different language the transcriber will translate the interview in the transcription process. We suggest this is done in a two-step process with transcription first and then translation. Note that when translating it becomes more important to let the participant read the translated interview.

We are only transcribing the open questions (currently, just Q4).

[1] Svend Brinkmann, Steinar Kvale: *InterViews: Learning the Craft of Qualitative Research Interviewing,* 3rd ed, SAGE, 2015.

3.1 Tools

You may transcribe using any software you are familiar with.

We have tested the web-service *otranscribe* (`http://otranscribe.com/`) that provides basic functionality. (start, stop, speed and editing of text).

We have also used *Express Scribe* (`http://www.nch.com.au/scribe/`), easy to use but we got some licensing problems.

3.2 Process Description

The following sections are adopted from Aine Humble's "Guide to Transcribing" (`http://www.msvu.ca/site/media/msvu/Transcription%20Guide.pdf`).

3.3 General formatting

All information that is not outspoken language and meta-information should be within square brackets "[ ]". The first line of each file should contain the following information, separated by semicolons:

1. Date and time for the interview.
2. Transcriber name.
3. Location of the interview.
4. Participant code.
5. Language the interview was carried out in.

Example:

```
[2017-05-05, 08:31; Daniel Toll; Linnaeus University, Sweden; DT1; Swedish]
I: Good morning. My name is Daniel Toll. Thank you for coming. [pause]
```

Transcripts should be saved in UTF8 format and named with the Participant code (e.g., DT1.txt) and saved within our shared repository here: http://share.somewhere.dot.

3.4 Time-stamps

If you want to include time-stamps, for example to locate particular parts in the transcription easily, you can include time-stamps. A time-stamps should indicate the time in hours, minutes, and seconds from the start of the recording ([time: HH:MM:SS]). The hours can be skipped when the transcription is shorter than one hour. Each time-stamp should be on a separate row.

Example:

```
[time: 03:14]
I: Please share an example of some actual code.
```

3.5 Speaker

Use "P:" for participant (interviewee) and "I:" for interviewer. These should always start on a new line and should be preceded by an empty line when there was a switch of "talker". You may use newlines within the text from a single talker.

Example:

I: Let's begin. Can you first tell me about what happened yesterday?

P: Sure, well the first thing I did was make an appointment with the doctor.

I: Ok.

[time: 04:15]
P: I had to wait on the phone for 30 minutes to talk to someone, it drove me crazy.
First I had to dial one number and choose from several options.

3.6 Pauses

Use ellipses (. . .) only at the end of sentences to indicate a "tailing off" of speaking or at the beginning of a person's statement if they are continuing from a previous thought and it might not make sense if you didn't indicate otherwise. Use [pause] only for long pauses.
For example:

P: He never did finish that up... I think he got tired.

3.7 Non-verbal communication

All non-verbal communication, such as pauses, laughter, crying, sighing, etc. should be identified with square brackets [ ] rather than parentheses ( ).
For example:

My father is so funny. [laughing]

Don't interpret the non-verbal communication.
For example, don't transcribe:

My father is funny. [nervous laughing]

Be consistent with how the same behavior is identified. For example, don't write [laughing] one time and [laughter] at another time.

3.8 Inaudible material

If particular words are inaudible, type [?] after the word(s) in parentheses rather than [can't make out]. If there are more than just one inaudible word or two, please mark this part with several question marks [???]. If you want to go back to an inaudible part, you can also include the time of the tape, e.g., [???: 05:16].

3.9 Quotations

If someone indicates that they said something to another person, indicate this with double quotations and use proper punctuation.
For example:

```
Then I said to him, “You shouldn’t do that”.
```

rather than:

```
Then I said to him you shouldn’t do that.
```

Similarly, do this for when people are indicating they were thinking something, such as: I was thinking, “Do I want to do this or not?”

3.10 Words that are emphasized

Capitalize words that the participant emphasizes in their talk (this will happen occasionally, but usually not too often).
For example:

```
I REALLY loved my wedding.
```

3.11 Confidentiality

Always recheck that identifying information such as all names are removed (name of the participant, any people they refer to, towns where they live, hospitals that they visit, countries they go on holiday to, etc.) To replace the specific word that you have removed, use a suitable acronym or pseudonym. Leaving empty spaces is not recommended, if it could be interesting for the analysis when such information is repeated.
For example, when the participant talks about her friend Sally she was visiting at the Bellevue Hospital Center:

```
I went to see S to see how she was doing.
OR
I went to the hospital to see how S was doing.

rather than
I went to the Bellevue Hospital Center to see how Sally was doing.
```

3.12 Proof reading

Transcripts always need to be reviewed! You will probably need to check through a transcript more than once.

Identify words or phrases that previously were thought to be inaudible and check for errors in the transcription (e.g., “can” instead of “can’t”, or a real name still in the transcript and thus needing to be removed). Listen very carefully to the tape and read the transcript carefully at the same time to locate these.

You should also make sure that you were consistent in coding non-verbal communication, pauses, etc.

3.13 Miscellaneous

When a person trails off on a word (doesn't finish it), or changes his/her thoughts part way through a sentence, type a long dash afterwards (this is called an "em-dash" (—)). Note that this is not the same as a hyphen (-). If producing an em-dash is cumbersome, we suggest using three hyphens in sequence instead.
For example:

To comp−−−to compensate for him not being there. rather than To comp, to compensate for him not being there. I went to—there was nobody there! rather than I went to, there was nobody there!

3.14 Refering to code

In order to be able to interpret Q4, where actual code-examples are discussed, we need to be able to capture the connection to the actual piece of code that was discussed. To do so, we include references to the code into the transcript. When the interviewer or interviewee mentions a code-section we just transcribe what was said, but if we have screen recording we want to include the code references in the following way.
For example when file HTTPReader.java is viewed/discussed and line 17 is pointed out particularly (e.g., by pointing with the mouse or finger at it):

[View: HTTPReader.java, Line: 4–30] I: This code is so horrible. [pause] Especially, this part. [laughing] [Select: HTTPReader.java, Line: 17]

3.15 Notes summary

Do we want/need a table to summarize all words in [ ] we want to use for easy reference?

- Start of transcript: [2017-05-05, 08:31; Daniel Toll; Linnaeus university, Sweden; DT1; Swedish]
- Code view change: [View: HTTPReader.java, Line: 4–30]
- Code selected : [Select: HTTPReader.java, Line: 17]
- Laughter: [laughing]
- Does not finish word: comp−−−
- Tailing off: He never finish that up... I think he got tired
- Longer break: [pause]
- Confidentiality: Remove names
- Emphasize: CAPITALIZE!
- Quotes: “You shouldn’t do that”
- Inaudible short: [?]
- Inaudible longer: [???]
- Inaudible longer with time: [???: 05:16]
- Time: [time: 04:15]
- Participant: P: What is this interview about?
- Interviewer: I: Let’s see...