

SELF Transcription Guide

To transcribe, you need at least:

- *A quiet environment, time and patience*
- *Good headphones*
- *MS Word (or equivalent)*
- *One of the following programmes: SoundScriber, Audacity, Praat (or equivalent)*

To download SoundScriber go to:

<http://www-personal.umich.edu/~ebreck/sscriber.html>

To download Audacity go to:

<http://audacity.sourceforge.net/>

To download Praat go to:

<http://www.fon.hum.uva.nl/praat/>

Transcription instructions

1. Event Descriptions (for courses)

At the beginning of each document, between angular brackets < >

NB! Angular brackets both at the beginning and at the end of the section

<FILE ID:

COURSE: (name of the course)

SESSION: (which session out of how many sessions, e.g. 2/7)

EVENT TYPE: (main event type, possible presentation titles etc.)

NOTES: (related material, other comments)

RECORDING DATE:

RECORDING DURATION:

RECORDING: (name(s) of those who recorded the event)

TRANSCRIBING:

RECHECKING:

PROOFREADING:

NUMBER OF PARTICIPANTS:

NUMBER OF SPEAKERS:

(Speakers with background information; NB! Each speaker has a specific code (e.g. S2) during the whole course so all transcriptions may not have S2, for instance.)

S1 = ORIGIN: (country); NATIVE LANGUAGE(S): (xx); OTHER LANGUAGES: (yy);

ACADEMIC ROLE: (title or position); GENDER: (zz); AGE: (choose from age group)

S2 = ... >

2. Transcription

- Write everything in small letters (except the exceptions below)
- Spelling: follow British English rules
- Speakers are marked with tags and numbered, i.e. <S1>, <S2> etc.
 - Utterance begins: <S1>
 - Utterance ends: </S1>
 - Unidentified speaker: <SU>
 - Uncertain speaker identification: <SU-1>
 - Several simultaneous speakers (usually laughter or sth): <SS>
- Uncertain transcription (replace “text” with what you hear): (text)
- Unintelligible speech: (xx)
- Laughter: @@
- Spoken laughing: @text@
- Pauses
 - Brief pause while speaking 2-3 sec.: ,
 - Pause 3-4 sec.: .
 - Pause 5 sec. or longer, rounded up to the nearest sec.: <P: 05>
 - example: <S1> this is er , an example </S1>
- Overlapping speech (approximate, shown to the nearest word, words not split by overlap tags):
 - <S1> [text] </S1>
 - <S2> [text text] text </S2>
- Backchannelling (during another speaker’ turn):
 - <S1> this is an example <S2> mhm </S2> you know </S1>
 - at least: okay, mhm-hm, mhm, uh-huh, uh-uh, yeah
- Hesitations
 - /öö/ → er
 - /(ö)m/ → erm
 - /aa/ (surprise) → ah
- Unfinished utterances: unfinis-
- Numbers as numbers: 10,000, 1932, 16, except those smaller than 10 (two or three, the second time, etc.) and decimals (0.6, 1.5)
- Names of *participants*: <NAME S1>
- Nonsense words: <SIC> text </SIC>
- Capital letters in all well-known expressions such as: TV, AM, PM, EU, USA

- Otherwise spelling out a word or acronym as letters: T-U-C, V-W
- Reading aloud: <READING> text </READING>
- Switching into another language than English (specify language if possible):
<FOREIGN> text </FOREIGN>
- Other events that affect the interpretation or comprehension of what is being said:
 - <WHISPERING>
 - <BACKGROUND NOISE>
 - <WRITING ON BLACKBOARD>
- Coughing, sighing, gasping, etc., if the speaker coughs etc. while speaking and this affects the situation or flow of speech (but NOT if other participants cough or sneeze, etc):
 - <COUGH>
 - <GASP>
 - <SIGH>
- Presentation starts or ends
 - <PRESENTATION BY S7 STARTS AT CA 6:16>
 - <PRESENTATION BY S7 ENDS AT CA 15:15>

3. Example

<FILE ID: V8D4Sp

COURSE: Agriculture and forestry in developing countries

SESSION: 6 of 7

EVENT TYPE: Seminar, student presentations by S5: Coconut and S6: Sustainable agriculture

NOTES: Presentation reports available, PP available from S6 only

RECORDING DATE: 26 Feb 2008

RECORDING DURATION: 1h 34min 31sec

RECORDING: Niina Hynninen & Pirjo Surakka-Cooper

TRANSCRIBING: Niina Hynninen

RECHECKING:

PROOFREADING:

NUMBER OF PARTICIPANTS: 12

NUMBER OF SPEAKERS: 8

T1: ORIGIN: Finland; NATIVE LANGUAGE(S): Finnish; OTHER LANGUAGES: English, Swedish, Spanish; ACADEMIC ROLE: University lecturer; GENDER: male; AGE: 45-60;

[...]

S6: ORIGIN: Spain; NATIVE LANGUAGE(S): Spanish; OTHER LANGUAGES: English, French; ACADEMIC ROLE: MA student; GENDER: male; AGE: 24-30;

[...]

S10: ORIGIN: Portugal; NATIVE LANGUAGE(S): Portuguese; OTHER LANGUAGES: English, Spanish; ACADEMIC ROLE: MA student; GENDER: female; AGE: 24-30 >

[...]

<S10> may i </S10>

<T1> yes <FIRST NAME S10> please </T1>

<S10> mhm do they have any er plan for the water treatment like anything that's going on like </S10>

<S6> no [(to be for a-)] </S6>

<S10> [no] no studies in that </S10>

<S6> for agriculture for drinking </S6>

<S10> (f- dr-) for er actually for drinking [because (it's)] </S10>

<S6> [for drinking no we don't] </S6>

<S10> the the worst problem [no] </S10>

<S6> [the] the water for er the drinking water the i- the er it doesn't kill them but it really hurts them <S10> mhm </S10> they all have the- these brown teeth er (problems) stomach problems but i think this is one of the smallest problems right now (so that) they it's really will be really really expensive there is not er at least i don't know any studies and i really familiarise with the- with this so i [guess] <S10> [mhm-hm] </S10> the- there is no at least for this place maybe there is so of course there is research in the world about this but not not related to saharawi and i know that for agriculture they (tried) to precipitate the phosphates before they they irrigated places but it's so so expensive that it's <S10> mhm </S10> worth [to] <S10> [mhm] </S10> just er continue with that (with) watering the plants and just leave the take the salt out of the soil when these white spots appear and that's it because it's no- it's not worth it's much money and all this is voluntary help there's nothing from the government there's nothing of that so so er i i go there and work no money (so just) </S6>

<S10> mhm-hm </S10>

<P: 06>

[...]