

Automatic transcription is not neutral

Wyke Stommel, Tom Koole, Tessa van Charldorp,
Sandra van Dulmen en Antal van den Bosch

ADVANT



rijksuniversiteit
groningen

Radboud Universiteit



Universiteit Utrecht

ADVANT

Automated annotation and analysis.



Tom Koole
Wyke Stommel
Tessa van Charldorp
Antal van den Bosch
Sandra van Dulmen



Automatic and manual annotation/ transcription

The image shows a video player interface with a family eating at a table. The video is annotated with colored bounding boxes: a yellow box around the mother, a green box around the boy in the middle, a blue box around the boy in the foreground, and a red box around the father. To the right of the red box, there are handwritten notes: "nn=", "on=", and "v=" with a small red square below them. Below the video is a transcription area with several horizontal bars, each containing a wavy line representing audio. On the left side of the video player, there is a vertical toolbar with various icons for video control.

Transcript in conversation analysis (CA)

- › Transcripts and video examined in conjunction
- › Transcription conventions developed over 45 years > Jefferson
- › Including pauses, overlap, intonation, breathing (in and out), clicks, laughing, crying etc.
- › Increasingly embodied behavior in transcript (Mondada)

Example CA transcript

01 Nancy: = I don't know it sounds kinda cra:zy =

02 Hyla: = 'hh [hhhh] =

03 Nancy: [bu: t] =

04 Hyla: = Jista liddle.

05 Nancy: We: : : ll,

06 (0.3)

07 Nancy: 'e may me feel bet[ter anywa (h) y] =

08 Hyla: [nhhhhhhhhhhhhh] =

09 (Hyla): = 'hk 'hhhhh

10 (0.4)

11 Nancy: So: .

12 (.)

13 Nancy: W[hat time,] eh hnh] =

Technology for the benefit of transcription?

- › Accelerates the process (speech recognition, image recognition?)
- › Possible to work with large(r) corpora
- › Objective measurement (e.g., silences)

Technology is not neutral

- 1) Theory-driven
- 2) Shows restricted set of aspects of interaction
- 3) Steers research questions/agenda's

Ochs 1979

Transcription is theory-driven:

- Transcriptions are the researcher's data
- Transcription is a selective process reflecting theoretical goals and definitions

=> Automatic transcription: theory and technology driven

- 1A. [NB-Assassination1:00:01:30:AUTO]
69 oh god long week
70 oh my god
71 i've decided sober i want you to have a t. v.
72
73 i won't either
73.5 (0.7)
74 like uh you know (0.1) that's where they
75 we took off on our charter flight that same spot
76 did you see it
77 (0.8)
78 and they took him and here uh you
79 know i wouldn't
80 watch it
81 i think it's so ridiculous i mean it's (0.4) it's a horrible
82 thing but my god (0.1) play up that's thing it's it's (.)
83 horrible
84 die people that
84.5 (0.3)
85 why is it a native american people think well they're no good
85.5 (0.5)
86 well they aren't very good some of
- 1B. [NB-Assassination1:00:01:30:JEFFERSON]
69 Lot: Oh: ↓Go:d a lo:ng wee[k. Yeah.]
70 Emm: [O h : my]↓God
71 I'm (.) glad it's over I won't eventurn the teevee
72 o[n.
73 Lot: [I won' eether.
74 Emm: °aOh no. They drag it out so° THAT'S WHERE THEY
75 WE TOOK OFF on ar chartered flight that sa:me spot
76 didju see it'
77 (0.7)
78 Emm: ·hh when they took him in[the airpla:ne,]
79 Lot: [n : N o : : . .] Hell I wouldn' ev'n
80 wa:tch it.
81 Lot: I think it's so ridiculous. I mean it's ·hhh it's a hôrrible
82 thing but my: Go:d. play up that thing it it's jst
83 ↑hôrri[ble.]
84 Emm: [It'II] drive people nu:ts.
85 Lot: Why id i-en makes Americ'n people think why ther no goo:d.
86 Emm: °°Mm:°° Well they aren't very good some of'm,

IBM Attila
speech
recognition:
“poor” audio
from the
60ies
(Moore
2015)

1A. [NB-Assassination1:00:01:30:AUTO]
 69 oh god long week
 70 oh my god
 71 i've decided sober i want you to have a t. v.
 72
 73 i won't either
 73.5 (0.7)
 74 like uh you know (0.1) that's where they
 75 we took off on our charter flight that same spot
 76 did you see it
 77 (0.8)
 78 and they took him and here uh you
 79 know i wouldn't
 80 watch it
 81 i think it's so ridiculous i mean it's (0.4) it's a horrible
 82 thing but my god (0.1) play up that's thing it's it's (.)
 83 horrible
 84 die people that
 84.5 (0.3)
 85 why is it a native american people think well they're no good
 85.5 (0.5)
 86 well they aren't very good some of

1B. [NB-Assassination1:00:01:30:JEFFERSON]
 69 Lot: Oh: ↓Go:d a lo:ng wee[k. Yeah.]
 70 Emm: [O h : my]↓God
 71 I'm (.) glad it's over I won't even turn the teevee
 72 o[n.
 73 Lot: [I won' eether.
 74 Emm: °aOh no. They drag it out so° THAT'S WHERE THEY
 75 WE TOOK OFF on ar chartered flight that sa:me spot
 76 didju see it/
 77 (0.7)
 78 Emm: ·hh when they took him in[the airpla:ne,]
 79 Lot: [n : N o : : : .] Hell I wouldn' ev'n
 80 wa:tch it.
 81 Lot: I think it's so ridiculous. I mean it's ·hhh it's a hōrrible
 82 thing but my: Go:d. play up that thing it it's jst
 83 ↑hōrri[ble.]
 84 Emm: [It'II] drive people nu:ts.
 85 Lot: Why id i-en makes Americ'n people think why ther no goo:d.
 86 Emm: °°Mm:°° Well they aren't very good some of'm,

Technology for transcription

Steers research questions/agendas (Bolden 2015)

- Favouring work on high quality recordings
- Favouring “text search” RQs (lexical items, discourse markers) as opposed to overlap, phonetic aspects, silences, etc.

Not neutral but useful?

- › Bolden 2015: “Going from an automatically produced transcript—with its missing speaker identifications, arbitrary line segmentation, word identification errors, etc.—to even a simple orthographic transcript where these shortcomings are corrected appears to be a very time-consuming task, without the analytic payoffs of the careful listening required for producing a CA transcript. It is, of course, possible that future versions of this software will address some of these problems and make automated transcription more cost effective. “



rijksuniversiteit
groningen

Radboud Universiteit



Universiteit Utrecht

ADVANT

Automated annotation and analysis.



Tom Koole
Wyke Stommel
Tessa van Charldorp
Antal van den Bosch
Sandra van Dulmen

Thank you